

21177

SEARCH REQUEST FORM

Examiner # (Mandatory): 69507 Requester's Full Name: John YLArt Unit 1646 Location (Bldg/Room#): CM 10815 Phone (circle 305 306 308) 4008Serial Number: 09/371333 Results Format Preferred (circle): PAPER DISK E-MAIL

Title of Invention _____

Inventors (please provide full names): _____

Earliest Priority Date: _____

Keywords (include any known synonyms registry numbers, explanation of initialisms): _____

Search Topic:

Please write detailed statement of the search topic, and the concept of the invention. Describe as specifically as possible the subject matter to be searched. Define any terms that may have a special meaning. Give examples of relevant citations, authors, etc., if known. You may include a copy of the abstract and the broadcast or most relevant claim(s).

please search SFO 712 10: 1, 2
and 7 of 09/371333.

FOR OFFICIAL USE ONLY

STAFF USE ONLY

Searcher: HART Type of Search _____ Vendors (include cost where applicable) _____

Searcher Phone #: <u>305-9203</u>	<u>1</u> N.A. Sequence	_____ STN
Searcher Location: <u>12C14</u>	<u>1</u> A.A. Sequence	_____ Questel/Orbit
Date Picked Up: <u>10/28/99</u>	_____ Structure (#)	_____ Lexis/Nexis
Date Completed: <u>11/1/99</u>	_____ Bibliographic	_____ WWW/Internet
Clerical Prep Time: <u>5</u>	_____ Litigation1	<u>X</u> In-house sequence systems (list)
Terminal Time: <u>5</u>	_____ Fulltext	_____ Dialog
Number of Databases: <u>11</u>	_____ Procurement	_____ Dr. Link
	_____ Other	_____ Westlaw
		_____ Other (specify)

COMPUGEN

BEST AVAILABLE COPY

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 29, 1999, 00:28:26 ; Search time 12.22 Seconds
(without alignments)
1262.292 Million cell updates/sec

Title: US-09-371-333-2
Perfect score: 1991
Sequence: 1 MWGRLLWPLVLGSLSGGT.....SKASAEGRGGMTHSSLLQ 385

Scoring table: BLOSUM62

Searched: 122810 seqs, 40065486 residues

Database : PIR_60:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	530	26.6	399	2 148705	proteinase activat
2	509	25.6	397	2 S66518	proteinase-activat
3	508	25.5	369	2 G02131	proteinase-activat
4	503	25.3	397	2 S64709	proteinase-activat
5	488	24.5	420	2 I51667	thrombin receptor
6	473	23.8	425	2 A37912	thrombin receptor
7	456.5	22.9	432	2 A43448	thrombin receptor
8	452	22.7	427	2 S17148	alpha-thrombin rec
9	372.5	18.7	308	2 I50241	G protein-coupled
10	371	18.6	326	2 G02514	P2 purinoceptor -
11	371	18.6	328	2 JC4800	P26 receptor - hu
12	367.5	18.5	328	2 I55450	a novel G protein-
13	365	18.3	350	1 A37963	complement C5a ana
14	345.5	17.4	370	2 JC5549	heptahelical P2Y5-
15	343	17.2	380	2 I38435	angiotensin recept
16	337.5	17.0	361	2 B45680	G protein-coupled
17	332.5	16.7	388	2 JN0605	somatostatin recep
18	332.5	16.7	355	2 JQ1231	interleukin-8 rece
19	326.5	16.4	350	2 A39445	interleukin-8 rece
20	326.5	16.4	353	2 JC2492	G protein-coupled
21	325	16.3	365	2 S68679	G protein-coupled
22	325	16.3	351	2 B42009	FMRF-related recep
23	323.5	16.2	418	2 A46226	somatostatin recep
24	322	16.2	364	2 JN0763	somatostatin recep
25	321.5	16.1	359	2 JC5277	G protein-coupled
26	320.5	16.1	360	2 A53611	interleukin-8 rece
27	320.5	16.1	363	2 I57955	somatostatin recep
28	320	16.1	373	2 JC4162	P2Y receptor - bov
29	320	16.1	363	2 I57940	somatostatin recep
30	320	16.1	387	3 JC5949	galanin receptor 2
31	319.5	16.0	428	2 A44021	somatostatin recep
32	318.5	16.0	351	2 A46525	complement C5a ana
33	317	15.9	355	2 G02436	C-C chemokine rece
34	317	15.9	355	2 A57237	chemokine (C-C) re
35	316.5	15.9	384	2 A47249	brain-specific som
36	316	15.9	373	2 JC4737	P2Y1 receptor G-pr
37	315	15.8	359	2 A42656	angiotensin II rec
38	314.5	15.8	355	2 JC5067	G protein-coupled
39	314.5	15.8	428	2 S30508	probable G protein

40 312.5 15.7 359 2 A48857 AT1 angiotensin II
41 312 15.7 362 2 S33733 G protein-coupled
42 311.5 15.6 359 2 S15403 angiotensin II rec
43 310 15.6 359 2 A48921 interleukin-8 rece
44 309.5 15.5 352 2 S27357 complement C5a ana
45 309.5 15.5 359 2 JH0621 angiotensin recept

ALIGNMENTS

RESULT 1

I48705
C:Species: Mus musculus (house mouse)
C>Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 02-Jul-1996
C:Accession: I48705
R:Nystedt, S.; Larsson, A.K.; Aberg, H.; Sundelin, J.
J. Biol. Chem. 270, 5950-5955, 1995
A>Title: The mouse proteinase-activated receptor-2 cDNA and gene. Molecular cloning a
A:Reference number: I48705; MUID:95197620
A:Accession: I48705
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-399 <RES>
A:Cross-references: EMBL:Z48043; NID:9663020; PID:9663021

Query Match 26.6% Score 530; DB 2; Length 399;
Best Local Similarity 34.1%; Pred. No. 3.7e-34;
Matches 131; Conservative 68; Mismatches 141; Indels 44; Gaps 11;

QY 13 GFSLSGGTQT-PSYDSESGTGGDDSTPSILPAPRGYPGQVCANDSDTLELPDSSRALL 71
DB 37 GRSLIGRLTQPTI-----TGKG-----VPVEPGF-----SIDEFSASIL 71
QY 72 LGWVTRRLVPALYGLVLPANGALWV-LATQAPRLPSTMLLMNATADILLALALP 130
DB 72 TGRKTTTFLPVYIIVFVIGLPSNGMALWIFLFTKKKHPAVIYMANLADLLSVWFP 131
QY 131 PRIAYHGRQWPPEGAACRLATAALYGHMGSVLLAAVSLDRYALVHPL---RARAL 187
DB 132 LKISYHLGNWVYGEALCKVLIGFFYGNMYCSILEFTCLSVQRYWVIVNPMGPRKA- 191
QY 188 RGRRLALGLCMAAWLMAALPLTLQRTQRLARSDRVLCHDALPLDAQASHWQPAFTC 247
DB 191 ---NIAVGSVLAIIWLLIFLVTIPLYVMKQTIYPALNITTTCHDVLPEVLVGMFNFLS 247
QY 248 LALLGCFL-PLIAMLICYGATLHTLAASG-----RRYGHALRLTAVVLASAVAFVPSN 300
DB 248 LA-IGVFLFPALLPASAYVLMIKTLRSSAMDEHSEKKRQRAIRLIITVLAIFYCFAPSN 306
QY 301 LLLLHSDPSAWGNLYGAYVPSLALSTLNSCVDPFIYVYVSAEFRDKVRAGLFORS- 360
DB 307 LLLVHVFLIKTQRQSHVYALYVALCLSLNSCIDPFYVYFVSKDFRDHARNALLCRSV 366
QY 360 -PGDTVASKASAEGRGGMTHSS 382
DB 367 RTVNRMQISLSSNFKSRKSGSYSS 390

RESULT 2

S66518
proteinase-activated receptor 2 precursor - human
C:Species: Homo sapiens (man)
C>Date: 28-Oct-1996 #sequence_revision 13-Mar-1997 #text_change 17-Mar-1999
C:Accession: S66518
R:Nystedt, S.; Emilsson, K.; Larsson, A.K.; Stroembeck, B.; Sundelin, J.
Eur. J. Biochem. 232, 84-89, 1995
A>Title: Molecular cloning and functional expression of the gene encoding the human p
A:Reference number: S66518; MUID:96048032
A:Accession: S66518
A>Status: preliminary

A:Molecule type: DNA
A:Residues: 1-397 <NYS>
A:Cross-references: EMBL:249993; NID:g1008084; PID:g1008085
C:Genetics:
A:Map position: 5q13
A:Introns: 28/1
F:1-36/Domain: activation peptide #status predicted <APT>
F:37-397/Product: proteinase-activated receptor 2 #status predicted <MAT>

Query Match 25.6%; Score 509; DB 2; Length 397;
Best Local Similarity 33.5%; Pred. No. 1.6e-32;
Matches 120; Conservative 64; Mismatches 156; Indels 18; Gaps 7;
QY 10 LVLGFSLSGTOPSPSYDDESGTGGDDSTPSILPAPRGVPGQVCANDSDTLELPDSSRA 69
DB 16 LAASLSCSTGIQTNRSSKGRSLIGKVDGTSV--TGKGVTVET-----VFSVDEFSAS 67
QY 70 LLLGWYPTRLVPALYGLVLYVGLPANGALWV-LATQAPRLPSTMLLMNLATADLLALA 128
DB 68 VLTGKLTTFVLPVIVTVFVVGLPNGMALWVLFRTKKKHPAVIYMANLADLLSVIW 127
QY 129 LPRIAYHLRGQWPGEACRLATLAALGHTGVTGSVLLAAVSLDRYLALVHPLRARLR 188
DB 128 FPLKIAHYHGNWNYGEALCNVLIFFGYGNMYCSILFMTCLSVQRYVWVIVNPM-GHSRK 186
QY 189 GRRALGLCMAAWLMAAALALPLTLQROTFLARSDRVLCCHDALPLDAQASHWQPAFTCL 248
DB 187 KANIAIGISLAIWLLILLVITPIYVVKQTFIPALNITTCDDVLPQLLVGDMFNFLSL 246
QY 249 ALGCFE-PLLAMLLCYGATLHTLAAG-----RRYGHALRLTAVVLAASAVAFVPSNL 301
DB 247 A-IGVFLPFAFLTASAYVLMIRLSSAMDENSEKKRRAIKLIVTVLAMYLICFTPSNL 305
QY 302 LLLLHSDSPSANGNLYGAYVPSLALSTLNSCDVDFIYYVSAEFRDKVRAGLQFORS 359
DB 306 LLVVHYFLIKSQGSHVYALYIVALCLSTLNSCIDPFYVYFVSHDFRDAKNAALLCRS 363

RESULT 3
G02131
proteinase-activated receptor 2 - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 21-Dec-1996 #sequence_revision 06-Jun-1997 #text_change 17-Jul-1998
C:Accession: G02131
R:Kahn, M.L.; Coughlin, S.R.
submitted to the EMBL Data Library, September 1995
A:Reference number: H00822
A:Accession: G02131
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-369 <KAH>
A:Cross-references: EMBL:U36753; NID:g1208539; PID:g1208540

Query Match 25.5%; Score 508; DB 2; Length 369;
Best Local Similarity 36.5%; Pred. No. 1.7e-32;
Matches 110; Conservative 57; Mismatches 124; Indels 10; Gaps 5;
QY 67 SRALLGWPTRLVPALYGLVLYVGLPANGALWV-LATQAPRLPSTMLLMNLATADLLL 125
DB 37 SASVLGKLTTFVLPVIVTVFVVGLPNGMALWVLFRTKKKHPAVIYMANLADLLS 96
QY 126 ALALPRIAYHLRGQWPGEACRLATLAALGHTGVTGSVLLAAVSLDRYLALVHPLRARLR 185
DB 97 VIWFLKIAHYHGNWNYGEALCNVLIFFGYGNMYCSILFMTCLSVQRYVWVIVNPM-GH 155
QY 186 ALRGRALGLCMAAWLMAAALALPLTLQROTFLARSDRVLCCHDALPLDAQASHWQPAF 245
DB 156 SRKANIAIGISLAIWLLILLVITPIYVVKQTFIPALNITTCDDVLPQLLVGDMFNFL 215
QY 246 TCLALGCFE-PLLAMLLCYGATLHTLAAG-----RRYGHALRLTAVVLAASAVAFVPS 298

DB 216 LSIA-IGVFLPFAFLTASAYVLMIRLSSAMDENSEKKRRAIKLIVTVLAMYLICFTP 274
QY 299 SNLLLLLHYSDPSANGNLYGAYVPSLALSTLNSCDVDFIYYVSAEFRDKVRAGLQFORS 358
DB 275 SNLLLLVHYFLIKSQGSHVYALYIVALCLSTLNSCIDPFYVYFVSHDFRDAKNAALLCR 334
QY 359 S 359
DB 335 S 335
RESULT 4
S64709
proteinase-activated receptor-2 precursor - human
C:Species: Homo sapiens (man)
C:Date: 14-Feb-1997 #sequence_revision 13-Mar-1997 #text_change 13-Nov-1998
C:Accession: S64709
R:Boehm, J. 314, 1009-1016, 1996
A:Title: Molecular cloning, expression and potential functions of the human proteinase
A:Reference number: S64709
A:Accession: S64709
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-397 <BOE>
A:Cross-references: EMBL:U34038; NID:g1041728; PID:g1041729
A:Note: the authors translated the codon GTC for residue 68 as Ile and AAC for residue
F:1-25/Domain: signal sequence #status predicted <SIG>
F:26-397/Product: proteinase-activated receptor-2 #status predicted <MAT>

Query Match 25.3%; Score 503; DB 2; Length 397;
Best Local Similarity 33.2%; Pred. No. 4.6e-32;
Matches 119; Conservative 64; Mismatches 157; Indels 18; Gaps 7;
QY 10 LVLGFSLSGTOPSPSYDDESGTGGDDSTPSILPAPRGVPGQVCANDSDTLELPDSSRA 69
DB 16 LAASLSCSTGIQTNRSSKGRSLIGKVDGTSV--TGKGVTVET-----VFSVDEFSAS 67
QY 70 LLLGWYPTRLVPALYGLVLYVGLPANGALWV-LATQAPRLPSTMLLMNLATADLLALA 128
DB 68 VLTGKLTTFVLPVIVTVFVVGLPNGMALWVLFRTKKKHPAVIYMANLADLLSVIW 127
QY 129 LPRIAYHLRGQWPGEACRLATLAALGHTGVTGSVLLAAVSLDRYLALVHPLRARLR 188
DB 128 FPLKIAHYHGNWNYGEALCNVLIFFGYGNMYCSILFMTCLSVQRYVWVIVNPM-GHSRK 186
QY 189 GRRALGLCMAAWLMAAALALPLTLQROTFLARSDRVLCCHDALPLDAQASHWQPAFTCL 248
DB 187 KANIAIGISLAIWLLILLVITPIYVVKQTFIPALNITTCDDVLPQLLVGDMFNFLSL 246
QY 249 ALGCFE-PLLAMLLCYGATLHTLAAG-----RRYGHALRLTAVVLAASAVAFVPSNL 301
DB 247 A-IGVFLPFAFLTASAYVLMIRLSSAMDENSEKKRRAIKLIVTVLAMYLICFTPSNL 305
QY 302 LLLLHSDSPSANGNLYGAYVPSLALSTLNSCDVDFIYYVSAEFRDKVRAGLQFORS 359
DB 306 LLVVHYFLIKSQGSHVYALYIVALCLSTLNSCIDPFYVYFVSHDFRDAKNAALLCRS 363

RESULT 5
I51667
thrombin receptor - African clawed frog
C:Species: Xenopus laevis (African clawed frog)
C:Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 13-Sep-1996
C:Accession: I51667
R:Gerzsten, R.E.; Chen, J.; Ishii, M.; Ishii, K.; Nanevici, T.; Turck, C.W.; Vu, T.H.
Nature 368, 648-651, 1994
A:Title: Thrombin receptor's specificity for agonist peptide is determined by its ext
A:Reference number: I51667; MUID:94195429
A:Accession: I51667
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA

A:Cross-references: EMBL:U09632; NID:g495197; PID:g495198

QY 156 LYGHMYGSVLLAAVSLDRYLALVHPLRARALRGRRRLALGLCMAAWLMAAALALPLTLQR 215

Db 410 LMPSKMDTCSSL 422

410 LMPSKMDTCSSHL 422

RESULT 8
S17148
alpha-thrombin receptor - Chinese hamster
C:Species: Cricetulus griseus (Chinese hamster)
C:Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 10-Sep-1997
C:Accession: S17148
R:B:Rasmussen, U.B.; Vouret-Craviari, V.; Jallat, S.; Schlesinger, Y.; Pages, G.; Pavirani
FEBs Lett. 288, 123-128, 1991
A:Title: cDNA cloning and expression of a hamster alpha-thrombin receptor coupled to Ca(2+)
A:Reference number: S17148; MUID:91348247
A:Accession: S17148
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-427 <RAS>
A:Cross-references: EMBL:X61958; NID:g940495; PID:g49538
C:Keywords: G protein-coupled receptor; transmembrane protein

```

Query Match      22.7% ; Score 452; DB 2; Length 427;
Best Local Similarity 32.1%; Pred. No. 4.6e-28;
Matches 125; Conservative 69; Mismatches 135; Indels 60; Gaps 14;
```


Qy	4	RLLWPLVLGSLSG---GTPTSPVYDDEGGST-----GGGD 36 : : : :
Db	5	RLLP--VAAGLUSLCGLSSRPVRQPSEMTDATNPRSFRLNPGETFELIPLGDEE 62 : : : :
Qy	37	DSTPSILPAPRGY-----PCQCANDSDTLELPDSSRALILGWVPTRL-VPAPLYGLVL 88 : : : : :
Db	63	EKNESTLPEGRAIYLKNSHPAIPFISE----DASGYLTSPWL--RLIPSYYTFVF 115 : : : : :
Qy	89	VVGPLPANGLAWLVIATQAP-RLPSTMILMNLIATADLLLALAPPRIATHLRGQRPPGEA 147 : : : : : : :
Db	116	VVSPLMLAIATAVFLKMKVKPAAVVYMLHAMADVLSVPLPKISIFYSGSQDWFGSG 175 : : : : : : :
Qy	148	ACRLATAALYGHMYSGLVLLAAVSLDRYLALVHLPHLARARALGRRLALGLCHAAWLMAAAL 207 : : : : : :
Db	176	MCRTAATAFYCNMYASIMLTIVSIDRELVAVVPIQSLSWTLGRANTCLIVWVMATMG 235 : : : : : :
Qy	208	ALPTLQAQTTRFLARSRVLCHDALP----LDAQASHWQPAFTCIALLGCFTPLPLAMLCL 264 : : : : : : :
Db	236	VWPLLKEQHTRPGLNITTCHDVNLNETLLQGFSYIFSAFSAFFL---VPLIISTICY 292 : : : : : : :
Qy	265	GATLHTLTAASG----RRYGHALRLTAVVLASAVAFFVPSNLLLLHY---SDPSPSANGN 317 : : : : : :
Db	293	MSITRCLSSSVANSRKSRALFSAAVCFVICFGFNVLIMHLYLLLDSP-ATEK 350 : : : : : :
Qy	318	LYGAYVPSLATLNSCVDPFIYYTVSAE 346 : : : : :
Db	351	AYFAYLLCVCVSYSSCCIDPLIIYYIASSE 379 : : : : :

```

RESULT      9
150241      G protein-coupled receptor 6H1 - chicken
N:Alternate names: purinoceptor 6H1
C:Species: Gallus gallus (Chicken)
C:Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 04-Sep-1998
C:Accession: 150241; JC4618
J:Immunol. 151, 628-636, 1993
R:Kaplan, M.H.; Smith, D.I.; Sundick, R.S.
A:Title: Identification of a G protein coupled receptor induced in activated
A:Reference number: 150241; MUID:93329058
A:Accession: 150241
A:Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: mRNA
A:Residues: 1-308 <KAP>
A:Cross-references: GB:L06109; NID:g304383; PID:g304384
R:Webb, T.E.; Kaplan, M.G.; Barnard, E.A.
Biochem. Biophys. Res. Commun. 219, 105-110, 1996
A:Title: Identification of 6H1 as a P2x purinoceptor: P2Y5.
A:Reference number: JC4618; MUID:96190677
A:Accession: JC4618
A:Molecule type: mRNA

```

A;Residues: 1-308 <WEB>
A;Cross-references: DDBJ:D49712
A;Experimental source: T-cells
C;Comment: This receptor plays a role in T-cell activation.
C;Genetics:
A;Gene: p2Y5
C;Superfamily: ATP receptor P2u
C;Keywords: G protein-coupled receptor; transmembrane protein
F:15-40/Domain: transmembrane #status predicted <TM1>
F:51-74/Domain: transmembrane #status predicted <TM2>
F:89-109/Domain: transmembrane #status predicted <TM3>
F:133-153/Domain: transmembrane #status predicted <TM4>
F:177-201/Domain: transmembrane #status predicted <TM5>
F:227-248/Domain: transmembrane #status predicted <TM6>
F:269-292/Domain: transmembrane #status predicted <TM7>

	Query Match	18.7%;	Score 372.5;	DB 2;	Length 308;
	Best Local Similarity	30.3%;	Pred. No. 5e-22;		
	Matches 91;	Conservative 57;	Mismatches 129;	Indels 23;	Gaps
Qy	58	SDTLELPDSSRALLLGWVTRLPV	ALYGLVVLVGLPANGIALKWL	A-TQAPRLPSTMLLM	116
Db	3	SSNCSTEDSKFTLYGCV-----	FSMVFLVGLIANCVAIYFTFLKVRNETTYML	54	
Qy	117	NLATADILLALAPPRITAYHLRG	QWPFGEAACLATAALYGHMVG	SVLLLAASVLDRYL	176
Db	55	NLAISDLLVFTLPFRITFYFV	VNNWPFQDVLCIKISVTLE	FTNMYSILFTLCISVDREL	113
Qy	177	ALVHPLRARLRRLGALGCGFL	PLAMLCYGAHTLTL---	AASGRYGHALRLTAVVLASAV	293
Db	114	AIVHPSKSLTRKRNARIVCVA	WITVLASTPASFEQSTNQNT	EQRTCFENPEST	173
Qy	237	QASHWQPAFTCLALLGCGFL	PLAMLCYGAHTLTL---	AASGRYGHALRLTAVVLASAV	293
Db	174	WKTYSLRIVFIEIVFGFPI	LIINLTCTMVTLRNKLPTLS	RNKLSSKKVLMKMFVHLV	233
Qy	294	AF---FVPSNLLILHYSDPS	AWGN-----LYGAYVPS	LALSTLNSCVDPFIYVYVS	344
Db	234	IFCFEFPYNITLIL-YS	LMRTQTWNCVSVTAVRTM	YPVTICIAVSNCCFDP	IVYFIS 292
RESULT	10				
G02514					
P2 purinoceptor - human					
C;Species: Homo sapiens (man)					
C;Date: 21-Dec-1996					
#sequence_revision 06-Jun-1997					
#text_change 04-Sep-1999					
C;Accession: G02514					
R;Hammer, F.; Southey, M.C.;					
Somers, G.R.;					
Hutchins, A.M.;					
Venter, D.J.					
submitted to the EMBL data Library,					
March 1996					
A;Reference number: H01373					
A;Accession: G02514					
A;Status: preliminary;					
translated from GB/EMBL/DDBJ					
A;Molecule type: mRNA					
A;Residues: 1-326 <HAM>					
A;Cross-references: EMBL:U52464;					
NID:g1407632; PID:g1407633					
C;Genetics:					
A;Gene: p2Y6					
C;Superfamily: ATP receptor P2u					

	Query Match	18.6%	Score 371;	DB 2;	Length 326;
	Best Local Similarity	31.4%;	Pred. No. 6.9e-22;		
	Matches 97; Conservative	46;	Mismatches 138;	Indels 28;	Gaps 7;
Qy	65 DSSRALLLGWPTR-----LVPALYGLVIVGILPANGLAWLWATQAQLPST-M 113	:	:	:	:
	: : :	: : :	: : :	: : :	: : :
Dd	2 DNGTGOALGLPPTCTCVRENFKQLLLPPVSAVLAAGLPNLCIVTOICTSRRLTRFVV 61	:	:	:	:
Qy	114 LLMNLATADLLLALAPRIAYHIGRGWPPEACACRLATAALYGHMYGVLLAAVASLD 173	:	:	:	:
	: : : : : :	: : : : : :	: : : : : :	: : : : : :	: : : : : :
Dd	62 YTLNALADLLYACSLPLLIYNAAQQDHPWPGDFACRLVRFYFANLHGHSILFTTCISFQ 121	:	:	:	:

biochemistry 32, 1243-1250, 1993
A>Title: Human chemotaxis receptor genes cluster at 19q13.3-13.4. Characterization of

A:Reference number: I52417; MUID:93192225
A:Accession: I52417
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-3 <RES>
A:Cross-references: GB:S56556; GB:S56557; NID:g298577; NID:g298578
C:Genetics:
A:Gene: GDB:C5R1; C5A; C5AR
A:Cross-references: GDB:128856; OMIM:113995
A:Map position: 19q13.3-19q13.4
A:Introns: 1/3
A>Note: the list of introns may be incomplete
C:Function:
A:Description: mediates the inflammatory and chemotactic responses of polymorphonuclear neutrophils
C:Superfamily: vertebrate rhodopsin
C:Keywords: chemotaxis; G protein-coupled receptor; glycoprotein; inflammation; polymorphonuclear neutrophil
F:1-37/Domain: extracellular #status predicted <EX1>
F:38-61/Domain: transmembrane #status predicted <TM1>
F:62-71/Domain: intracellular #status predicted <IN1>
F:72-94/Domain: transmembrane #status predicted <TM2>
F:95-110/Domain: extracellular #status predicted <EX2>
F:111-132/Domain: transmembrane #status predicted <TM3>
F:133-149/Domain: intracellular #status predicted <IN2>
F:150-174/Domain: transmembrane #status predicted <TM4>
F:175-206/Domain: extracellular #status predicted <EX3>
F:207-227/Domain: transmembrane #status predicted <TM5>
F:228-242/Domain: intracellular #status predicted <IN3>
F:243-264/Domain: transmembrane #status predicted <TM6>
F:265-283/Domain: extracellular #status predicted <EX4>
F:284-307/Domain: transmembrane #status predicted <TM7>
F:308-350/Domain: intracellular #status predicted <IN4>
F:5/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 18.3%; Score 365; DB 1; Length 350;
Best Local Similarity 29.9%; Pred. No. 2.2e-21;
Matches 104; Conservative 58; Mismatches 140; Indels 46; Gaps 9;

QY 56 NDSDTLEL----PDSSRALLGWPTPLVPALYGLVVLVGLPANGALAVLATOAPRLPS 111
DB 15 DDKDTLDLNTVPDKTNTLR--VPDILALVIFAVVFLVGLVGLNVALVWVTFEAKRTIN 71
QY 112 TMLLMNLATADLLALAPRIAYHLRGQWPGEACRLATAALYGHMYSVLLAAVS 171
DB 72 AIWFLNLAVALDFSLALPILFTSIQVHHWPEGGAACSLPSILLNMYAAILLATIS 131
QY 172 LDRYLALVPLRLARLGRRLALGLCMALMAALALPLTLQRTFLARSD----RVL 227
DB 132 ADRFLVFPKPCQNPFRGAGLAWIACAVAW---GLALLTIPSEFLYRVREYFPKVL 187
QY 228 CHDALPLDAQASHWQPAFTCLAL---LCGFLPLAMLLCYG-ATLHTLAASGRRYGHAL 282
DB 188 C-----GVDSHDKRRERAVLVLVGLFLPLLTTCYTFILLRTWSRATRTKTL 241
QY 283 RLTAVVLASAVFAFFVSNLL-LLLHYSDSPSAMGNLYGAYVPSLALSTNSCDPFIY 341
DB 242 KVVAVVAVSFFIFWLPYQVTGIMMSFLEPSPPTFLNKLKLSLCSVFAINCINPIIV 301
QY 342 YVSAERDKVRAGL-----FQSPGDTVASKASA 370
DB 302 VAGQFGQGRKRLKSLPLLRNLVTEESVVRKSKSFTRSTVDTMAQKTOA 349

RESULT 14
JC5549
heptahelical P2Y5-like receptor - human
C:Species: Homo sapiens (man)
C:Date: 02-Sep-1997 #sequence_revision 05-Sep-1997 #text_change 17-Mar-1999
A:Accession: JC5549
R:Janssens, R.; Boeynaems, J.M.; Godart, M.; Communi, D.
Biochem. Biophys. Res. Commun. 236, 106-112, 1997
A:Title: Cloning of a human heptahelical receptor closely related to the P2Y5 receptor.

A:Reference number: JC5549; MUID:97366605
A:Accession: JC5549
A:Molecule type: DNA
A:Residues: 1-370 <JAN>
A:Cross-references: DDBJ:AF005419; NID:g2240034; PID:g2240035
C:Superfamily: ATP receptor P2u

Query Match 17.4%; Score 345.5; DB 2; Length 370;
Best Local Similarity 29.2%; Pred. No. 7.7e-20;
Matches 95; Conservative 61; Mismatches 138; Indels 31; Gaps 9;

QY 37 DSTPSILPAPRGYQGVCAQNDSTLELPSRALLGWPTPLVPALYGLVVLVGLPANG 96
DB 13 DNSSELRPR---LGNATAN--NTCIVDDSFKNLNG-----AVYSVFLGLITNS 58
QY 97 ZALWLYATQAP-RLPSTMLLMNLATADLLALAPRIAYHLRGQWPGEACRLATAA 155
DB 59 VSLFVFCFRMKMRSETAIFITNLAVSDLLFVCTLPFKIFYNF-NRHWPFEGDTLCKISGTA 117
QY 156 LYGHMYSVLLAAVSIDRYALVPLRLARLGRRLALGLCMALMAALALPLTLQRT 215
DB 118 FUTNIYGSMLFTCISVDRLAIVYFPRSTIRTRNSAIVCAGVWILVLSGGISASLFS 177
QY 216 QTFRLARSDRVLCHDALPLDAQASHWQPAFTCLALLGFLPLAMLLCYGATLH----- 270
DB 178 TT--NVNNTATTCFEGLSKRVWKYLSKITIFIEVVGFIPLILNVSCSVVLTLRKPA 235
QY 270 TLAASGRRYGHALRTAVVLASAVFAFFVPSNLLLLHYSDSPSAMGNLY-----GAYV 323
DB 236 TLSQIGTNNKKVKLMTITVMAVVFVFCFYPYNSVFLF-YALVRSQAITNCFLEFAKIMYP 294
QY 324 PSLALSTNSCDPFIYVVSAREP 348
DB 295 ITCLATLNLCCDFPIFYITLESFQ 319

RESULT 15
I38435
angiotensin receptor homolog APJ - human
C:Species: Homo sapiens (man)
C:Date: 15-Mar-1996 #sequence_revision 15-Mar-1996 #text_change 15-Mar-1996
C:Accession: I38435
R:O'Dowd, B.F.; Heiber, M.; Chan, A.; Heng, H.H.; Tsui, L.; Kennedy, J.L.; Shi, X.; P.
Gene 136, 355-360, 1993
A:Title: A human gene that shows identity with the gene encoding the angiotensin receptor
A:Reference number: I38435; MUID:94124031
A:Accession: I38435
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-380 <RES>
A:Cross-references: EMBL:U03642; NID:g425351; PID:g425352
C:Genetics:
A:Gene: APJ
A:Map position: 11q12
A:Introns: #status absent

Query Match 17.2%; Score 343; DB 2; Length 380;
Best Local Similarity 29.5%; Pred. No. 1.2e-19;
Matches 97; Conservative 58; Mismatches 142; Indels 32; Gaps 9;

QY 79 LVPALYGLVVLVGLPANGALAVL--ATQAPRLPSTMLLMNLATADLLALAPRIAYH 136
DB 30 LIPALYMLVFLITGTGGLVLTWTVFRSSRRKRSADFIASLAVADLTFFVTLPLWATYT 89
QY 137 LRQWPGEACRLATAALYGHMYSVLLAAVSIDRYALVPLRLARLGRRLALGL 196
DB 90 YRDYDWPFGTFCKLSLYLIEFNMYASVFCGLSFDRLYLAIRPVANARLRLRVSGAVA 149
QY 197 CMAWLMMAALALPLTLQRTFLARSDRVLCHDALPLDAQASH---WQPAF-TCLALLG 252
DB 150 TAVLWVLAALLAMPVWVLRITGDLNTTKVQCYNMDYSMATVATVSEWAVEVGLVGSSTTVG 209

Qy 253 CFLPLAMLCYGATLHTLAASGRR-----YGHALRLTAVVVLASAVAF---FVPSNLLLL 304
Db 210 FWPFTIMLTCTCYFFFAQTIAGHFRKERIEGLRKRRLLSIIIVLVVTFALCWMFYHLVKT 269
Qy 305 LHYSDDPSANGNLYGAVVPSL-----ALSTLNSCVDPFIIYYVSAEPRDKVR----- 353
Db 270 L-YMLGSLHLWPCDDFLFMNIFPYCTCISYVNSCLNPFYAFEDPRFRQACTSMLCCGQ 328
Qy 353 ---AGLFQSPGDTVASKASAEAGGSRGMG 378
Db 329 SRCAGTSHSSSGEKSASYSS--GHSQGP 355

Search completed: October 29, 1999, 02:35:48
Job time: 7642 sec

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 29, 1999, 11:07:23 ; Search time 11.26 Seconds
(without alignments)
966.545 Million cell updates/sec

Title: US-09-371-333-2
Perfect score: 1991
Sequence: 1 MWGRLLNPLVLGFSLSGGT.....SKASABGGSGMGTHSSLLQ 385

Scoring table: BLOSUM62

Searched: 77977 seqs, 28268293 residues

Database : SwissProt_37.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	530	26.6	399	1	PAR2_MOUSE
2	524	26.3	397	1	PAR2_RAT
3	521.5	26.2	374	1	PAR3_HUMAN
4	509	25.6	397	1	PAR2_HUMAN
5	507.5	25.5	369	1	PAR3_MOUSE
6	488	24.5	420	1	THRR_XENLA
7	480	24.1	425	1	THRR_PAPHA
8	473	23.8	425	1	THRR_HUMAN
9	468.5	23.5	430	1	THRR_MOUSE
10	458.5	23.0	428	1	THRR_CRILLO
11	456.5	22.9	432	1	THRR_RAT
12	372.5	18.7	308	1	P2Y5_CHICK
13	371	18.6	328	1	P2Y6_HUMAN
14	367.5	18.5	328	1	P2Y6_RAT
15	366	18.4	340	1	C5AR_MACMU
16	365	18.3	350	1	C5AR_HUMAN
17	362.5	18.2	340	1	C5AR_PONPY
18	362	18.2	340	1	C5AR_GORGO
19	362	18.2	340	1	C5AR_PANTR
20	354.5	17.8	339	1	GPRH_HUMAN
21	346	17.4	328	1	P2Y3_CHICK
22	345.5	17.4	352	1	C5AR_RAT
23	343	17.2	380	1	APJ_HUMAN
24	341.5	17.2	370	1	P2Y2_HUMAN
25	339.5	17.1	360	1	IL8B_BOVIN
26	337.5	17.0	361	1	EBI2_HUMAN
27	332.5	16.7	355	1	IL8A_RABIT
28	332.5	16.7	388	1	SSR4_HUMAN
29	331.5	16.6	344	1	P2Y5_HUMAN
30	329.5	16.5	350	1	IL8A_PANTR
31	328	16.5	348	1	FML1_GORGO
32	328	16.5	537	1	P2Y8_XENLA
33	326.5	16.4	353	1	FML1_RAT
34	325	16.3	351	1	FML1_HUMAN
35	325	16.3	365	1	P2Y4_HUMAN
36	323.5	16.2	350	1	IL8A_HUMAN
37	323.5	16.2	418	1	SSR3_HUMAN
38	323	16.2	357	1	CG96_HUMAN
39	320.5	16.1	360	1	IL8B_HUMAN
40	320.5	16.1	363	1	SSR5_HUMAN
41	320	16.1	353	1	IL8B_PANTR
42	320	16.1	373	1	P2YR_BOVIN
43	320	16.1	363	1	SSR5_RAT

44	319.5	16.0	428	1	SSR3_MOUSE	P30935 mus musculu
45	319	16.0	360	1	GPRP_HUMAN	O00155 homo sapien
ALIGNMENTS						
RESULT 1	PAR2_MOUSE	STANDARD;	PRT;	399 AA.		
ID	PAR2_MOUSE	STANDARD;	PRT;	399 AA.		
AC	P55086;					
DT	01-OCT-1996 (REL. 34, CREATED)					
DT	01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)					
DT	01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)					
DE	PROTEINASE ACTIVATED RECEPTOR 2 PRECURSOR (PAR-2).					
GN	PAR2 OR GPR11 OR GPCR11.					
OS	MUS MUSCULUS (MOUSE).					
OC	EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;					
OC	RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; MUS.					
RN	[1]					
RP	SEQUENCE FROM N.A.					
RX	MEDLINE; 95197620.					
RA	NYSTEDT S., LARSSON A.-K., AABERG H., SUNDELIN J.;					
RT	"The mouse proteinase-activated receptor-2 cDNA and gene. Molecular cloning and functional expression.";					
RT	J. BIOL. CHEM. 270:5950-5955(1995).					
RL	-1- FUNCTION: ORPHAN RECEPTOR.					
CC	-1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.					
CC	-1- PTM: ACTIVATED BY PROTEOLYTIC CLEAVAGE OF ITS EXTRACELLULAR N-TERMINUS.					
CC	-1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.					
CC	-----					
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).					
CC	-----					
DR	EMBL; Z48043; G663021; -					
DR	GCRDB; GCR_1727; -					
DR	MGD; MGI:101910; GPCR11.					
DR	PROSITE; PS00237; G-PROTEIN_RECEPTOR; FALSE_NEG.					
DR	PFAM; PF00001; 7tm_1; 1.					
DR	HSSP; P34996; 1DDD.					
KW	G-PROTEIN COUPLED RECEPTOR; TRANSMEMBRANE; GLYCOPROTEIN; SIGNAL.					
FT	SIGNAL	1	25	POTENTIAL.		
FT	PROPEP	26	38	REMOVED FOR RECEPTOR ACTIVATION (BY SIMILARITY).		
FT	CHAIN	39	399	PROTEINASE ACTIVATED RECEPTOR 2.		
FT	DOMAIN	39	77	EXTRACELLULAR (POTENTIAL).		
FT	TRANSNEM	78	103	1 (POTENTIAL).		
FT	DOMAIN	104	112	CYTOPLASMIC (POTENTIAL).		
FT	TRANSNEM	113	132	2 (POTENTIAL).		
FT	DOMAIN	133	151	EXTRACELLULAR (POTENTIAL).		
FT	TRANSNEM	152	173	3 (POTENTIAL).		
FT	DOMAIN	174	192	CYTOPLASMIC (POTENTIAL).		
FT	TRANSNEM	193	213	4 (POTENTIAL).		
FT	DOMAIN	214	243	EXTRACELLULAR (POTENTIAL).		
FT	TRANSNEM	244	262	5 (POTENTIAL).		
FT	DOMAIN	263	287	CYTOPLASMIC (POTENTIAL).		
FT	TRANSNEM	288	310	6 (POTENTIAL).		
FT	DOMAIN	311	325	EXTRACELLULAR (POTENTIAL).		
FT	TRANSNEM	326	349	7 (POTENTIAL).		
FT	DOMAIN	350	399	CYTOPLASMIC (POTENTIAL).		
FT	DISULFID	150	228	BY SIMILARITY.		
FT	CARBOHYD	33	33	POTENTIAL.		
FT	CARBOHYD	224	224	POTENTIAL.		
SQ	SEQUENCE	399 AA;	44752 MW;	D4EA5805 CRC32;		

Query Match

26.6%; Score 530; DB 1; Length 399;

FT	PROPEP	26	36	REMOVED FOR RECEPTOR ACTIVATION (BY SIMILARITY).
FT	CHAIN	37	397	PROTEINASE ACTIVATED RECEPTOR 2.
FT	DOMAIN	37	75	EXTRACELLULAR (POTENTIAL).
FT	TRANSNEM	76	101	1 (POTENTIAL).
FT	DOMAIN	102	110	CYTOPLASMIC (POTENTIAL).
FT	TRANSNEM	111	130	2 (POTENTIAL).
FT	DOMAIN	131	149	EXTRACELLULAR (POTENTIAL).
FT	TRANSNEM	150	171	3 (POTENTIAL).
FT	DOMAIN	172	190	CYTOPLASMIC (POTENTIAL).
FT	TRANSNEM	191	211	4 (POTENTIAL).
FT	DOMAIN	212	241	EXTRACELLULAR (POTENTIAL).
FT	TRANSNEM	242	260	5 (POTENTIAL).
FT	DOMAIN	261	285	CYTOPLASMIC (POTENTIAL).
FT	TRANSNEM	286	308	6 (POTENTIAL).
FT	DOMAIN	309	323	EXTRACELLULAR (POTENTIAL).
FT	TRANSNEM	324	347	7 (POTENTIAL).
FT	DOMAIN	348	397	CYTOPLASMIC (POTENTIAL).
FT	DISULFID	148	226	BY SIMILARITY.
FT	CARBOHYD	23	23	POTENTIAL.
FT	CARBOHYD	222	222	POTENTIAL.
SEQ	SEQUENCE	397 AA;	44440 MW; D59C4FEC CRC32;	
Query Match 26.3%; Score 524; DB 1; Length 397;				
Best Local Similarity 34.6%; Pred. NO. 2.4e-27;				
Matches 124; Conservative 62; Mismatches 132; Indels 40; Gaps 9;				
QY	13	GFSLSGGTQTPSVYDESGSTGGDDSTPSTLPAPRGYPGOVCANDSDTLELPDSSRALLL	72	
DB	35	GRSLIGRLDTPPI-----TGKG-----APVEPGF-----SVDEFSASVLT	70	
QY	73	GWVPTRLVPALYGLVLVVGLPANGLAWLWLTQAPRL--PSTMLLMNLATDLLLLALALPP	131	
DB	71	KLTVLPVYIIIVFVIGLPSNGMALWVFFTKKHPAVIYMANLADLSVWFPL	130	
QY	132	RIAYHLRGQRPFGEEAACRLATAALYGHMYGSVLLLAASLDRLYLALVHPL---	188	RARALR
DB	131	KISYHLHGNDWTYGDALCKVLGFFGYGNMYCSILEMCTLSVQYVWVIVNPMGHSKRRA--	189	
QY	189	GRLALGLCHAAWLMMAALALPLTLOQTRFLARSDEVLCHDALPLDQAQASHQWPAFTCL	248	
DB	189	--NIAVGSLAIIWLLIFLVIPIYVVRQTIYIPALNTITTCVDLPEVLVGMFSYFLSL	246	
QY	249	ALLGCLF--PLLAMLLCYGATLHTLAASG-----RRYGHALRLTAVVLASAVAFVPSNL	301	
DB	247	A--IGVFLFPALLTASAVYLMIKILRSSAMDEHEKKRRRAIRLIITVLSMYFFCFAPS	305	
QY	302	LLLHYSDPSPSANGNLYGAYVPSLALSTLNSCVDPFIYIVYSAEFRDKVRAGLFRS	359	
DB	306	LLVHYFLIKSQSHVYALYLVALCLSTNSCIDPVEYVFSKDFRDQARNALLCRS	363	
RESULT 3				
PAR3_HUMAN	PAR3_HUMAN	STANDARD;	PRT;	374 AA.
ID	AC	O00254;		
DT	15-JUL-1998	(REL. 36, CREATED)		
DT	15-JUL-1998	(REL. 36, LAST SEQUENCE UPDATE)		
DE	15-JUL-1998	(REL. 36, LAST ANNOTATION UPDATE)		
DE	DE	PROTEINASE ACTIVATED RECEPTOR 3 PRECURSOR (PAR-3) (THROMBIN RECEPTOR		
DE	2).			
GN	GN	PAR3.		
OS	OS	HOMO SAPIENS (HUMAN).		
OC	OC	EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;		
OC	OC	PRIMATES; CATARRHINI; HOMINIDAE; HOMO.		
RN	RN	[1]		
RX	RX	SEQUENCE FROM N.A.		
RX	RX	MEDLINE: 97242411.		
RA	RA	ISHIHARA H., CONNOLLY A.J., ZENG D., KAHN M.L., ZHENG Y.W.,		
RA	RA	TIMMONS C., TRAM T.T., COUGHLIN S.R.;		
RT	RT	"Protease-activated receptor 3 is a second thrombin receptor in humans."		

RL NATURE 396:502-506(1997).
 CC -1- FUNCTION: RECEPTOR FOR ACTIVATED THROMBIN.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
 CC -1- PTM: IT IS THOUGHT THAT CLEAVAGE AFTER AA 38 BY THROMBIN LEADS TO
 CC ACTIVATION OF THE RECEPTOR. THE NEW AMINO TERMINUS FUNCTIONS AS
 CC A TETHERED LIGAND AND ACTIVATES THE RECEPTOR.
 CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; U92971; G1938375; -
 DR GCRDB; GCR_1320; -
 DR MIN; 601919; -
 DR PROSITE; PS00237; G_PROTEIN_RECEPTOR; 1.
 DR PFAM; PF00001; 7tm_1; 1.
 DR HSSP; P34996; 1DDD.
 DR G-PROTEIN COUPLED RECEPTOR; TRANSMEMBRANE; GLYCOPROTEIN; SIGNAL;
 KW BLOOD COAGULATION.
 FT SIGNAL 1 2
 FT PROPEP 1 38
 FT CHAIN 39 374
 FT DOMAIN 39 94
 FT TRANSMEM 95 120
 FT DOMAIN 121 128
 FT TRANSMEM 129 148
 FT DOMAIN 149 167
 FT TRANSMEM 168 189
 FT DOMAIN 190 206
 FT TRANSMEM 207 230
 FT DOMAIN 231 259
 FT TRANSMEM 261 280
 FT DOMAIN 281 297
 FT TRANSMEM 298 322
 FT DOMAIN 323 336
 FT TRANSMEM 337 361
 FT DOMAIN 362 374
 FT SITE 38 39
 FT DISULFID 166 245
 FT CARBOHYD 25 25
 FT CARBOHYD 82 82
 FT CARBOHYD 331 331
 FT SEQUENCE 374 AA; 42508 MW; B1794373 CRC32;
 Query Match 26.2%; Score 521.5; DB 1; Length 374;
 Best Local Similarity 33.8%; Pred. No. 3.4e-27;
 Matches 120; Conservative 62; Mismatches 158; Indels 15; Gaps 5;
 Qy 3 GRLLWPLVLFSLGGTQ-----TPSYDESSTGGGDDSTPSILPAPRGYQGV----- 54
 Db 10 GILLLLPTFCOSGMDNTNNAKPTLPKTRGAPPNSFEFFP--FSALEGWGTATITVK 67
 Qy 54 --CANDST-LLELPSSRALLLGWVPTRLVVALYGLVGLVGLPANGALWVLATQAPRLP 110
 Db 68 IKCPESASHLVKVNATMGYVTSLSUSTLPAIYLLVFGVGPANAVTLWMLFFTRTSIC 127
 Qy 111 STMLMNLATADLLALAPPRIAHYLRGQRPFGFGEACRLATATAALYGHMYGSVLLAAV 170
 Db 128 TTVFTNLAIVHDELFCVTLPEKIAHLNNGNWNVEVLCAATTVIFGNYMYSILLACI 187
 Qy 171 SLDRYLAIVHDELARALGRRLKIALGLCMANWMAAALPLTLQGRFRLARSORVLCHD 230
 Db 188 SINRYLAIVHPTFVTRGLPKHYVALVTCGLVWATVFLYMLPFILKQEYLVLPDITTCGD 247
 Qy 231 ALPLDAQASHNQ-PAFTCLALLGCFLLPALLMLCYGATLHTLAASGRYGHALRLTAVVL 289

Db 248 VHTCESSPFQLYYFISLAFFGLIPFVLIIYCYAAIHTLNAYDRWLVYKASLLIL 307
 Qy 290 ASAVAFVPSNLLLLHYSDPSFSGMNGLYGAYVPSLALSTLNSCLSDVPFIYVVS 344
 Db 308 VIFTICFAPSNIILIIHANYNNNTDGLYFVLIALLCLGSLNSCLDPFIFYELMS 362
 RESULT 4
 PAR2_HUMAN STANDARD; PRT; 397 AA.
 ID PAR2_HUMAN AC P5085; Q13317; Q13346;
 DT 01-OCT-1996 (REL. 34, CREATED)
 DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
 DT 13-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
 DE PROTEINASE ACTIVATED RECEPTOR 2 PRECURSOR (PAR-2).
 GN PAR2 OR GPR11.
 OS HOMO SAPIENS (HUMAN).
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
 OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 96048032.
 RA NYSTEDT S., EMLISSON K., LARSSON A.-K., STROEMBECK B., SUNDELIN J.;
 RT "Molecular cloning and functional expression of the gene encoding the
 RT human proteinase-activated receptor 2.";
 RL EUR. J. BIOCHEM. 232:84-89(1995).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 96177879.
 RA BOHM S., KONG W., BROMME D., SNEEKENS S.P., ANDERSON D.C.,
 RA PAYAN D.G., BURNETT N.W.;
 RT "Molecular cloning, expression and potential functions of the human
 RT proteinase-activated receptor-2.";
 RL BIOCHEM. J. 314:1009-1016(1996).
 RN [3]
 RP SEQUENCE OF 29-397 FROM N.A.
 RA KAHN M.L., COUGHLIN S.R.;
 RL SUBMITTED (SEP-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
 CC -1- FUNCTION: ORPHAN RECEPTOR.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
 CC -1- TISSUE SPECIFICITY: WIDELY EXPRESSED IN TISSUES WITH ESPECIALLY
 CC HIGH LEVELS IN PANCREAS, LIVER, KIDNEY, SMALL INTESTINE, AND
 CC COLON. MODERATE EXPRESSION IS DETECTED IN MANY ORGANS, BUT NONE IN
 CC BRAIN OR SKELETAL MUSCLE.
 CC -1- PTM: ACTIVATED BY PROTEOLYTIC CLEAVAGE OF ITS EXTRACELLULAR N-
 CC TERMINUS.
 CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; Z49993; G1008085; -
 DR EMBL; Z49994; G1008085; JOINED.
 DR EMBL; U34038; G1041729; -
 DR EMBL; U36753; G1208540; -
 DR GCRDB; GCR_1964; -
 DR MIN; 600933; -
 DR PROSITE; PS00237; G_PROTEIN_RECEPTOR; FALSE_NEG.
 DR PFAM; PF00001; 7tm_1; 1.
 DR HSSP; P34996; 1DDD
 KW G-PROTEIN COUPLED RECEPTOR; TRANSMEMBRANE; GLYCOPROTEIN; SIGNAL.
 FT SIGNAL 1 25
 FT PROPEP 26 36
 FT CHAIN 37 397
 FT DOMAIN 37 75
 FT TRANSMEM 76 101


```
RESULT 6
THRR_XENLA STANDARD; PRT; 420 AA.
AC P47749;
DT 01-FEB-1996 (REL. 33, CREATED)
DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE THROMBIN RECEPTOR PRECURSOR.
OS XENOPUS LAEVIS (AFRICAN CLAWED FROG).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; AMPHIBIA; BATRACHIA; ANURA;
OC MESOBATRACHIA; PIPODEA; PIPIDAE; XENOPODINAE; XENOPUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 94195429.
RA GERSTEN R.E., CHEN J., ISHII M., ISHII K., NANEVICZ T.,
RA TURCK C.W., VU T.-K.H., COUGHLIN S.R.;
RT "Specificity of the thrombin receptor for agonist peptide is defined
by its extracellular surface."
RL NATURE 368:648-651(1994).
CC -1- FUNCTION: RECEPTOR FOR ACTIVATED THROMBIN.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- PTM: IT IS THOUGHT THAT CLEAVAGE AFTER AA 42 BY THROMBIN LEADS TO
ACTIVATION OF THE RECEPTOR. THE NEW AMINO TERMINUS FUNCTIONS AS
A TETHERED LIGAND AND ACTIVATES THE RECEPTOR.
CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC EMBL: U09632; G495198;
CC GCRDB: GCR_1038;
CC PROSITE; PS00237; G-PROTEIN_RECEPTOR; 1.
CC PFAM; PF00001; 7tm_1; 1.
CC G-PROTEIN COUPLED RECEPTOR; TRANSMEMBRANE; GLYCOPROTEIN; SIGNAL;
KW BLOOD COAGULATION.
FT SIGNAL 1 20 POTENTIAL.
FT PROPEP 21 42 REMOVED FOR RECEPTOR ACTIVATION.
FT CHAIN 43 420 THROMBIN RECEPTOR.
FT DOMAIN 43 101 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 102 127 1 (POTENTIAL).
FT DOMAIN 128 136 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 137 156 2 (POTENTIAL).
FT DOMAIN 157 175 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 176 197 3 (POTENTIAL).
FT DOMAIN 198 217 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 218 238 4 (POTENTIAL).
FT DOMAIN 239 267 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 268 287 5 (POTENTIAL).
FT DOMAIN 288 310 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 311 333 6 (POTENTIAL).
FT DOMAIN 334 345 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 346 369 7 (POTENTIAL).
FT DOMAIN 370 420 CYTOPLASMIC (POTENTIAL).
FT SITE 42 43 CLEAVAGE (BY THROMBIN).
FT DISULFID 174 253 BY SIMILARITY.
FT CARBOHYD 38 38 POTENTIAL.
FT CARBOHYD 86 86 POTENTIAL.
FT MUTAGEN 42 42 K-2A; LACK OF ACTIVATION BY THROMBIN.
SQ SEQUENCE 420 AA; 47435 MW; D21E8698 CRC32;

Query Match 24.5%; Score 488; DB 1; Length 420;
Best Local Similarity 34.9%; Pred. No. 5.4e-25;
Matches 124; Conservative 60; Mismatches 143; Indels 28; Gaps 8;
QY 27 DESGSGDDSTPSILPAPRGVQGVCAVNDSTLELPDSSRALLLGGWTRLPVLYGL 86
```

```
Db 63 DESGE-GSGDQA-----PVSRSARKFIRRNITKEAQYLSQ-----WL-TKVPSTLYTV 110
QY 87 VLVVGLPANGALWV-LAQAPRLPSTMLMNLATALLALALPPRIAYHLRGQWPF 145
Db 111 VFIVGLPLNLLAIIIFLEKMKVRKPAVVMYMLNLAIDVFFVSVLPFKIAYHLSGNDWLF 170
QY 146 EACRATATAALYGHMYGVSLLLAASLDRLYLALVHLRLARALGRRLALGLCHNAWMLAA 205
Db 171 PGMCRIVTAIFYCNMYCSVLLIASISVDRFLAVVYPMHSLSWRTMSRAYNACSFIMLISI 230
QY 206 ALALPTLQROTFLRARSDRVLCCHDALPDAQASHQWPAFTCLALLGCFPLLLAMLILCYG 265
Db 231 ASIPLLVTEQTKIPRLDTTCHVDLDLKDLDKDFIYFSSFCLLFFFFFIITICYI 290
QY 266 ATLHTLAAS-----GRRYGHALRLTAVVLASAVAFFVPSNLLLLHSDPSAMGNLYGA 321
Db 291 GIIRSLSSSIENCKKTRALFLAVVVLGVFIICFGPTNVLFTHVLQEADEF---LYFA 347
QY 322 YVPSLALSTUNSCVDPIFYIVYSAEFRDKVRAGLFORSPOGDTVASKASAEAGSGRG 376
Db 348 YILSACVSGVSCCLDPLIYYIYASSQCRVLYSLCCR-----KVSEPGSGSTG 394

RESULT 7
THRR_PAPHA STANDARD; PRT; 425 AA.
AC P56488;
DT 15-JUL-1998 (REL. 36, CREATED)
DT 15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE THROMBIN RECEPTOR PRECURSOR.
GN F2R OR PAR1 OR BTHR12.
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC PRIMATES; CATARRHINI; CERCOPIITHECIDAE; CERCOPIITHECINAE; PAPIO.
RN [1]
RP SEQUENCE FROM N.A.
RA SHOJI M., HAYZER D.J., HANSON S.R.;
RL SUBMITTED (NOV-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -1- FUNCTION: RECEPTOR FOR ACTIVATED THROMBIN.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- PTM: IT IS THOUGHT THAT CLEAVAGE AFTER AA 41 BY THROMBIN LEADS TO
ACTIVATION OF THE RECEPTOR. THE NEW AMINO TERMINUS FUNCTIONS AS
A TETHERED LIGAND AND ACTIVATES THE RECEPTOR.
CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC EMBL; AF028727; G3605877;
CC GCRDB; GCR_2523;
CC PROSITE; PS00237; G-PROTEIN_RECEPTOR; 1.
CC G-PROTEIN COUPLED RECEPTOR; TRANSMEMBRANE; GLYCOPROTEIN; SIGNAL;
KW BLOOD COAGULATION.
FT SIGNAL 1 26 POTENTIAL.
FT PROPEP 27 41 REMOVED FOR RECEPTOR ACTIVATION.
FT CHAIN 42 425 THROMBIN RECEPTOR.
FT DOMAIN 42 102 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 103 128 1 (POTENTIAL).
FT DOMAIN 129 137 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 138 157 2 (POTENTIAL).
FT DOMAIN 158 176 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 177 198 3 (POTENTIAL).
FT DOMAIN 199 218 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 219 239 4 (POTENTIAL).
FT DOMAIN 240 268 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 269 288 5 (POTENTIAL).
```

```
FT DOMAIN 289 311 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 312 334 6 (POTENTIAL).
FT DOMAIN 335 350 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 351 374 7 (POTENTIAL).
FT DOMAIN 375 425 CYTOPLASMIC (POTENTIAL).
FT CARBOHYD 35 35 POTENTIAL.
FT CARBOHYD 62 62 POTENTIAL.
FT CARBOHYD 75 75 POTENTIAL.
FT CARBOHYD 250 250 POTENTIAL.
FT CARBOHYD 259 259 POTENTIAL.
FT SITE 41 42 CLEAVAGE (BY THROMBIN).
FT DOMAIN 57 60 ASP/GLU-RICH (ACIDIC).
FT DISULFID 175 254 BY SIMILARITY.
SQ SEQUENCE 425 AA; 47253 MW; 9E709862 CRC32;

Query Match 24.1%; Score 480; DB 1; Length 425;
Best Local Similarity 33.9%; Pred. No. 1.8e-24;
Matches 117; Conservative 70; Mismatches 144; Indels 14; Gaps 5;

QY 53 VCANDSDTLEP-----DSSRALLGVVPTRLVPLVGLVLPANGALV-LAT 104
DB 72 VSINKSPLOKPLPAFISDAGYLTSSWL-TLFVPSVITGVFVSLPVNMAIVVFIK 130
QY 105 QAPRLPSTMLNLTADLLALPPRIAYHLRGORPFGAACLATAALYGHMYSV 164
DB 131 MKVKPVPVYMLHLATADVLSVLPFKISYLSGSDWQFSGELCFVTAAYFCNNYASI 190
QY 165 LLLAAVSLDRYALVPLRARRALGRRLALGICMAWMAALALPLTLQRTFLARS 224
DB 191 LLMTVISIDRFVAVVPMQSLSWRTLGASFTCLALWALAIAGVVPVLLKEOTIQV 250
QY 225 RVLCHDALPDAQASHWQPAFTCLALLGCFPLLAMLCYCATHTLA---ASGRYGH 280
DB 251 ITTCHDVLTETLEGYYAYFAFSAFNVFFVPLIISTVCVSIIRCLSSSTVANSKSR 310
QY 281 ALRLAVLASAVAFVPSNLLLLHYDPS-PSAWGNLYGAYVPSLALSTLNSCVDPE 339
DB 311 ALFLSAVFCFIICFGPTNILLIAHYSPLSTSTTEAAYFAYLLVCVSSISCCIDPL 370
QY 340 YYYVSAEPRDKVRAGLFQSPGDTVASKASAGSGRGMGTHSSLL 384
DB 371 YYYASQRYVYSILCKESSDPSNSSGQIMASKMDTCSNL 415

RESULT 8
ID THRR_HUMAN STANDARD; PRT: 425 AA.
AC P25116;
DT 01-MAY-1992 (REL. 22, CREATED)
DT 01-MAY-1992 (REL. 22, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE THROMBIN RECEPTOR PRECURSOR.
GN F2R OR PAR1 OR TR.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 91168254.
RA VU T.-K.H., HUNG D.T., WHEATON V.I., COUGHLIN S.R.;
RT "Molecular cloning of a functional thrombin receptor reveals a novel
RT proteolytic mechanism of receptor activation.";
RL CELL 64:1057-1068(1991).
CC -!- FUNCTION: RECEPTOR FOR ACTIVATED THROMBIN.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -!- TISSUE SPECIFICITY: PLATELETS AND VASCULAR ENDOTHELIAL CELLS.
CC -!- PTM: IT IS THOUGHT THAT CLEAVAGE AFTER AA 41 BY THROMBIN LEADS TO
CC ACTIVATION OF THE RECEPTOR. THE NEW AMINO TERMINUS FUNCTIONS AS
CC A TETHERED LIGAND AND ACTIVATES THE RECEPTOR.
CC -!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
```

```
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M62424; G339677; -
DR PIR; A37912; A37912.
DR GCRDB; GCR_0088; -
DR MIN; 187930; -
DR PROSITE; PS00237; G-PROTEIN_RECEPTOR; 1.
DR PFAM; PF00001; 7tm1.1.
KW G-PROTEIN COUPLED RECEPTOR; TRANSMEMBRANE; GLYCOPROTEIN; SIGNAL;
KW BLOOD COAGULATION.
FT SIGNAL 1 26 POTENTIAL.
FT PROPEP 27 41 REMOVED FOR RECEPTOR ACTIVATION.
FT CHAIN 42 425 THROMBIN RECEPTOR.
FT DOMAIN 42 102 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 103 128 1 (POTENTIAL).
FT DOMAIN 129 137 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 138 157 2 (POTENTIAL).
FT DOMAIN 158 176 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 177 198 3 (POTENTIAL).
FT DOMAIN 199 218 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 219 239 4 (POTENTIAL).
FT DOMAIN 240 268 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 269 288 5 (POTENTIAL).
FT DOMAIN 289 311 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 312 334 6 (POTENTIAL).
FT DOMAIN 335 350 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 351 374 7 (POTENTIAL).
FT DOMAIN 375 425 CYTOPLASMIC (POTENTIAL).
FT CARBOHYD 35 35 POTENTIAL.
FT CARBOHYD 62 62 POTENTIAL.
FT CARBOHYD 75 75 POTENTIAL.
FT CARBOHYD 250 250 POTENTIAL.
FT CARBOHYD 259 259 POTENTIAL.
FT SITE 41 42 CLEAVAGE (BY THROMBIN).
FT DOMAIN 57 60 ASP/GLU-RICH (ACIDIC).
FT DISULFID 175 254 BY SIMILARITY.
SQ SEQUENCE 425 AA; 47410 MW; E9A485AE CRC32;

Query Match 23.8%; Score 473; DB 1; Length 425;
Best Local Similarity 32.1%; Pred. No. 5e-24;
Matches 133; Conservative 76; Mismatches 169; Indels 36; Gaps 10;

QY 4 RLLMLPLVLFSLSG---GTQTPSVYDESGTGGDDSTFSLPAP-----RG 48
DB 5 RLLL--VAACFSLGCLLSARTRARRPESKATNATLDRSFLRNPNDKYEPFWEDEKN 62
QY 49 YPG-----QVCANDSDYL--ELP-----DSSRALLGVVPTRLVPLVGLVLPANG 96
DB 63 ESGLEYRLVSVINKSPLOKPLPAFISDAGYLTSSWL-TLFVPSVITGVFVSLPLNI 121
QY 97 LALWV-LATQAPRLPSTMLNLTADLLALPPRIAYHLRGORPFGAACLATAA 155
DB 122 MAIVVFILMKVKPVPVYMLHLATADVLSVLPFKISYFSGSDWQFSGELCFVTA 181
QY 156 LYGHMYSVLLAAVSLDRYALVPLRARRALGRRLALGICMAWMAALALPLTLQ 215
DB 182 FYCNMYASILLMTVISIDRFVAVVPMQSLSWRTLGASFTCLALWALAIAGVVPV 241
QY 216 QTFRLASDRVLCDALPDAQASHWQPAFTCLALLGCFPLLAMLCYCATHTLA 275
DB 242 QTIQVPLNITTDHVLNETLEGYYAYFAFSAFNVFFVPLIISTVCVSIIRCLSS 301
QY 276 ----RRYGHALRLTAVLASAVAFVPSNLLLLHYDPS-PSAWGNLYGAYVPSL 330
DB 302 VANSKSKRALFLSAVFCFIICFGPTNILLIAHYSPLSTSTTEAAYFAYLLVCV 361
QY 331 LNSCVDPPYYVSAEPRDKVRAGLFQSPGDTVASKASAGSGRGMGTHSSLL 384
```


DE THROMBIN RECEPTOR PRECURSOR.
GN F2R OR PAR1.
OS CRICETULUS LONGICAUDATUS (LONG-TAILED HAMSTER) (CHINESE HAMSTER).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC RODENTIA; SCIUROGNATHI; MURIDAE; CRICETINAE; CRICETULUS.
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE=LUNG;
RX MEDLINE: 91348247.
RA RASMUSSEN U B., VOURET-CRAVIARI V., JALLAT S., SCHLESINGER Y.,
RA PAGES G., PAVIRANI A., LECOQ J.P., FOURSEGUR J.,
RA OBERGHEEN-SCHILLING E.;
RT "cDNA cloning and expression of a hamster alpha-thrombin receptor
coupled to Ca2+ mobilization.";
RL FEBS LETT. 288:123-128(1991).
RN [2]
RP SEQUENCE OF 42-428 FROM N.A.
RX MEDLINE: 96028007.
RA HARTMANN T., GRACE M.B., BUZARD G.S., RUOSS S.J.;
RT "Thrombin receptor polymorphism in Chinese hamster.";
RL BIOCHEM. BIOPHYS. RES. COMMUN. 215:974-980(1995).
CC -1- FUNCTION: RECEPTOR FOR ACTIVATED THROMBIN.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- PTM: IT IS THOUGHT THAT CLEAVAGE AFTER AA 41 BY THROMBIN LEADS TO
ACTIVATION OF THE RECEPTOR. THE NEW AMINO TERMINUS FUNCTIONS AS
A TETHERED LIGAND AND ACTIVATES THE RECEPTOR.
CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC
CC EMBL: X61958; G49538; -.
CC DR EMBL: U34047; G1000286; -.
CC DR PIR: S17148; S17148.
CC DR GCRDB; GCR_0166; -.
CC DR GCRDB; GCR_0432; -.
CC DR GCRDB; GCR_1173; -.
CC DR PROSITE; PS00237; G_PROTEIN_RECEPTOR; 1.
CC DR PFAM; PF00001; 7tm_1; 1.
CC DR HSSP; P34996; 1DDO.
CC KW G-PROTEIN COUPLED RECEPTOR; TRANSMEMBRANE; GLYCOPROTEIN; SIGNAL;
KW BLOOD COAGULATION; POLYMORPHISM.
FT SIGNAL 1 ?
FT PROPEP 1 ?
FT CHAIN 42 428
FT DOMAIN 42 428
FT TRANSMEM 106 131
FT DOMAIN 132 140
FT TRANSMEM 141 160
FT DOMAIN 161 179
FT TRANSMEM 180 201
FT DOMAIN 202 221
FT TRANSMEM 222 242
FT DOMAIN 243 271
FT TRANSMEM 272 291
FT DOMAIN 292 314
FT TRANSMEM 315 337
FT DOMAIN 338 352
FT TRANSMEM 353 377
FT DOMAIN 378 428
FT SLICE 41 42
FT DOMAIN 60 63
FT DISULFID 178 257
FT CARBOHYD 65 65
FT CARBOHYD 78 78
FT CARBOHYD 253 253
FT CARBOHYD 262 262

FT VARIANT 84 84 MISSING (IN AN ALLELE).
FT CONFLICT 384 384 H -> T (IN REF. 1).
SQ SEQUENCE 428 AA; 47602 MW; BC709D4C CRC32;
Query Match 23.0%; Score 458.5; DB 1; Length 428;
Best Local Similarity 32.4%; Pred. No. 4.4e-23;
Matches 126; Conservative 69; Mismatches 135; Indels 59; Gaps 14;
Qy 4 RLLPLVPLGFSLSG---GTQTPSYVDESGST-----GGGD 36
Db 5 RLLL--VAAGSLGCLPILLSRVPRQPESEMTDATVNPFRSLRNPGEFTFELPLGDEE 62
Qy 37 DSTPSILPAPGYGQVCANDSDTELP-----DSSRALLGWVPTL--VPALYGLVL 88
Db 63 EKNESTLPEGRA---IYLNKSHPPAPLAFIFEDASGYLTSPWL--RLIPSYVTFV 116
Qy 89 VVGLPANGALWVLTQAP--RLPSTMLMLNLTADLLALALPPRIAYHLRGQRWPFGEA 147
Db 117 VVSLPLNLITAIATVFLKMKVKKPAVVYMLHMLAMADVLFVPLKISYFSGSDWQFCG 176
Qy 148 ACRLATAALGHMYGCVLLLAASVLDRLVALVHLRRLRGRRLALGLCAAWLMAAL 207
Db 177 MCRFATAFYCNMYASIMLMTVISIDRELVAVYPIQSLSWRTGLRANFTCLVIVWMAIM 236
Qy 208 ALPLTQRTFLRARSRLVCHDALP---LDAQASHMQPFTCLALLGCFPLLLAMLICY 264
Db 237 VVPLLLKEQTVTRVGLNITTCHDVNLQGFYSYFSAFSAVFFL---VPLIISTICY 293
Qy 265 GATLHTLAAG-----RRYGHALRTAVVLASAVAFFVPSNLLLLHY---SDPSPSANGN 317
Db 294 MSIIIRCLSSSVANRKRKRALFLSAAVFCVFCVGTNVLIMHYLLSD-SP-ATEK 351
Qy 318 LYGAVPSIALSTLNSCVDPETIYYVSAE 346
Db 352 AYFALLCVSVSSVSCCIDPLIYYIASSE 380
RESULT 11
THRR_RAT
ID THRR_RAT STANDARD; PRT; 432 AA.
AC P26824;
DT 01-AUG-1992 (REL. 23, CREATED)
DT 01-AUG-1992 (REL. 23, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE THROMBIN RECEPTOR PRECURSOR.
GN F2R OR PAR1.
OS RATULUS NORVEGICUS (RAT).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE=ARTICULAR SMOOTH MUSCLE;
RX MEDLINE: 92381002.
RA ZHONG C., HAYZER D.J., CORSEN M.A., WICK K., RUNGE M.S.;
RT "Molecular cloning of the rat vascular smooth muscle thrombin
receptor. Evidence for in vitro regulation by basic fibroblast growth
factor.";
RL J. BIOL. CHEM. 267:16975-16979(1992).
CC -1- FUNCTION: RECEPTOR FOR ACTIVATED THROMBIN.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- PTM: IT IS THOUGHT THAT CLEAVAGE AFTER AA 45 BY THROMBIN LEADS TO
ACTIVATION OF THE RECEPTOR. THE NEW AMINO TERMINUS FUNCTIONS AS
A TETHERED LIGAND AND ACTIVATES THE RECEPTOR.
CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC

```
CC EMBL; M81642; G207466; -.
DR PIR; A43448; A43448.
DR GCRDB; GCR_0263; -.
DR PROSITE; PS00237; G_PROTEIN_RECEPTOR; 1.
DR PFAM; PF00001; 7tm_1; 1.
DR HSP; P34996; 1DDD.
DR G-PROTEIN COUPLED RECEPTOR; TRANSMEMBRANE; GLYCOPROTEIN; SIGNAL;
KW BLOOD COAGULATION.
FT SIGNAL 1 ?
FT PROPEP ? 45
FT CHAIN 46 432
FT DOMAIN 46 109
FT TRANSNEM 110 135
FT DOMAIN 136 144
FT TRANSNEM 145 164
FT DOMAIN 165 183
FT TRANSNEM 184 205
FT DOMAIN 206 225
FT TRANSNEM 226 246
FT DOMAIN 247 275
FT TRANSNEM 276 295
FT DOMAIN 296 318
FT TRANSNEM 319 341
FT DOMAIN 342 357
FT TRANSNEM 358 381
FT DOMAIN 382 432
FT SITE 45 46
FT DOMAIN 64 68
FT DISULFID 182 261
FT CARBOHYD 69 82
FT CARBOHYD 82 257
FT CARBOHYD 257 266
FT CARBOHYD 266 266
FT SEQUENCE 432 AA; 48280 MW; 758B2DA8 CRC32;

Query Match 22.9%; Score 456.5; DB 1; Length 432;
Best Local Similarity 30.5%; Pred. No. 5.9e-23;
Matches 132; Conservative 79; Mismatches 155; Indels 67; Gaps 15;

QY 4 RLLWPLVGLFSLGSG---GTQTPSVYDESGTGGDDSTPSILPAPRGYQGVCANSD-D 59
DB 5 RLLW---VAVGLSLGCPILLSSRPVMPQPESEK---YATPYATPNRPF---FLRNPS 55
QY 60 TLE-----LPDSSRALLLG-----WVPTLVPA 82
DB 56 TFEQFPLGDEEKNESIPLEGRAVYLNKSRPPMPFPFISEDASGYLTSPWL-TLIPS 114
QY 83 LYGLVVLVGLPANGALWVLAQAP-RLPSTMLMNLATADLLALALPPRIAYHLRQOR 141
DB 115 VYTFVFLVSLPLNLAIAVFRMKVKPAAVVMHLAMADVLSVLPFKISYFSGTD 174
QY 142 WPFGEACRLATAALGHMYGVSLLLAASLDRLVALVHLPLARALGRRLALGLCAAW 201
DB 175 WQFSGMCRFAACYCNMYSIMLMTVISIDRELVAVVYIQSLSWRTLGRANFTCVVW 234
QY 202 LMAAALALPTLQRTFLARSDRVLCHDALP---LDAQASHMQPFTCLALLGCFPL 258
DB 235 VMAINGVVPILLKEQTQVPLNITTCHVLTLLHGFYSYFSAFSAIFL---VPLI 291
QY 259 AMLLCYGATLHTLAASG-----RRYGHALRLTAVVLASAVAFVPSNLLLLHY---SDPS 311
DB 292 ISTVCYTSIIRCLSSAVANRKSRAFLFLSAAVFCIFVFCGFTNVLIVHLLLS-D-S 350
QY 312 PSANGNLYGAVPSLALSTLNCVDYDFIYVYVSAEERDKVRAGLFQSPGDTKASAE 371
DB 351 PGT-ETAYFAYLLGCVTSVASCIDPLIYYVYASSECOKHLYSILCCRSDNSCNSTGQ 409
QY 372 GGSRGMTGTHSLL 384
DB 410 LMPKMDTCSSHL 422
```

```
RESULT 12
P2Y5_CHICK
ID P2Y5_CHICK STANDARD; PRT; 308 AA.
AC P32250;
DT 01-OCT-1993 (REL. 27, CREATED)
DT 01-OCT-1993 (REL. 27, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE P2Y PURINOCEPTOR 5 (P2Y5) (PURINERGIC RECEPTOR 5) (6H1).
GN P2Y5.
OS GALLUS GALLUS (CHICKEN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ARCHOSAURIA; AVES;
OC NEOGNATHAE; GALLIFORMES; PHASIANIDAE; PHASIANINAE; GALLUS.
[1]
RN RP SEQUENCE FROM N.A.
RC MEDLINE; 93329058.
RX KAPLAN M.H., SMITH D.I., SUNDICK R.S.;
RA "Identification of a G protein coupled receptor induced in activated
RT T cells."
RL J. IMMUNOL. 151:628-636(1993).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- TISSUE SPECIFICITY: INDUCED IN ACTIVATED T-CELLS.
CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; L06109; G304384; -.
DR GCRDB; GCR_0689; -.
DR PROSITE; PS00237; G_PROTEIN_RECEPTOR; 1.
DR PFAM; PF00001; 7tm_1; 1.
DR HSP; P34996; 1DDD.
KW G-PROTEIN COUPLED RECEPTOR; TRANSMEMBRANE; GLYCOPROTEIN; LIPOPROTEIN;
KW PALMITATE.
FT DOMAIN 1 16 EXTRACELLULAR (POTENTIAL).
FT TRANSNEM 17 43 1 (POTENTIAL).
FT DOMAIN 44 52 CYTOPLASMIC (POTENTIAL).
FT TRANSNEM 53 76 2 (POTENTIAL).
FT DOMAIN 77 89 EXTRACELLULAR (POTENTIAL).
FT TRANSNEM 90 109 3 (POTENTIAL).
FT DOMAIN 110 130 CYTOPLASMIC (POTENTIAL).
FT TRANSNEM 131 151 4 (POTENTIAL).
FT DOMAIN 152 178 EXTRACELLULAR (POTENTIAL).
FT TRANSNEM 179 206 5 (POTENTIAL).
FT DOMAIN 207 224 CYTOPLASMIC (POTENTIAL).
FT TRANSNEM 225 250 6 (POTENTIAL).
FT DOMAIN 251 269 EXTRACELLULAR (POTENTIAL).
FT TRANSNEM 270 289 7 (POTENTIAL).
FT DOMAIN 290 308 CYTOPLASMIC (POTENTIAL).
FT CARBOHYD 5 5 POTENTIAL.
FT LIPID 281 281 PALMITATE (BY SIMILARITY).
FT DISULFID 86 165 POTENTIAL.
FT SEQUENCE 308 AA; 35597 MW; B2DFEAD4 CRC32;

Query Match 18.7%; Score 372.5; DB 1; Length 308;
Best Local Similarity 30.3%; Pred. No. 1.1e-17;
Matches 91; Conservative 57; Mismatches 129; Indels 23; Gaps 7;

QY 58 SDILEPDSRRALLGHWPTPLPALLYGLVVLVGLPANGALWVLA-TQAPRLPSTMLLM 116
DB 3 SSNCSTEDSFYTLGCV-----FSMVFVLGLIANCAVAYIFTFTLKVNETTYML 54
QY 117 NLATADLLALALPPRIAYHLRQORWPFGEACRLATAALGHMYGVSLLLAASLDRL 176
DB 55 NLAISDLLFVTLFPRI-YFVYVNRWFFGDVLCKISTVLTFTYNNYGSILFCLISVDRLF 113
```


CC MEDIATED BY G PROTEINS WHICH ACTIVATE A PHOSPHATIDYLINOSITOL-
CC CALCIUM SECOND MESSENGER SYSTEM. FUNCTIONALLY COUPLED TO
CC PHOSPHOLIPASE C.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -!- TISSUE SPECIFICITY: ABUNDANTLY EXPRESSED IN VARIOUS TISSUES
CC INCLUDING LUNG, STOMACH, INTESTINE, SPLEEN, MESENTERY, HEART, AND,
CC MOST PROMINENTLY AORTA.
CC -!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: D63665; G1066008; -
CC GCRDB: GCR_1520; -
CC PROSITE: PS00237; G_PROTEIN_RECEPTOR; FALSE_NEG.
CC PFAM: PF00001; 7tm_1; 1.
CC HSP: P34996; 1DD.
CC G-PROTEIN COUPLED RECEPTOR; TRANSMEMBRANE; GLYCOPROTEIN.
KW DOMAIN 1 27 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 28 48 1 (POTENTIAL).
FT DOMAIN 49 62 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 63 83 2 (POTENTIAL).
FT DOMAIN 84 101 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 102 122 3 (POTENTIAL).
FT DOMAIN 123 144 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 145 165 4 (POTENTIAL).
FT DOMAIN 166 194 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 195 215 5 (POTENTIAL).
FT DOMAIN 216 236 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 237 257 6 (POTENTIAL).
FT DOMAIN 258 280 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 281 303 7 (POTENTIAL).
FT DOMAIN 304 328 EXTRACELLULAR (POTENTIAL).
FT DISULFID 99 177 BY SIMILARITY.
FT CARBOHYD 5 5 POTENTIAL.
SQ SEQUENCE 328 AA; 36677 MW; 2FA27ESC CRC32;

Query Match 18.5%; Score 367.5; DB 1; Length 328;
Best Local Similarity 31.1%; Pred. No. 2.5e-17;
Matches 103; Conservative 47; Mismatches 136; Indels 45; Gaps 10;
QY 57 DSDTLEPDSRALLLGWVPT- - - - -LVPALYGLVVLVGLPANGAL-WVLATQ 105
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
D 4 DNGTIQAP- - - - -GLPTTCVYREDFRLLPPVSVLVGLPLNVCVIAQICASR 55
QY 106 APRLPSTMLNMLNATADLLALALPPRIAYHLRGQRPFGEEAARLATAALYGHMYSVL 165
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
D 56 RLTLSRAVYTLNLADLLVACSLPLLIYVARGDHPFGDLACRLVFLFYANLHGSIL 115
QY 166 LLAASVLDRLVLAHPLARALR-GRRLALGLCMAWLMMAALALPLTLQRTFLARSD 224
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
D 116 FLTCISFQRYLGICHPLAPWPKHRRGRRAAWVGVVWLVVTAQCLPTAVFAATG--IQRN 173
QY 225 RVLCHDALPLDAOSHWPAFTCLALIGCFLLPILAMLCYCATLHTLA- - - - -ASGR 276
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
D 174 RTVCYDLSP-PIILSTRYLPYGMALTVIGFLPPTALLACVCMARLCRODGPAGVQAE 232
QY 277 RYGHALRLTAVLASAVFAVPSNL- - - - -LLLLHYSDPS- - - - -AWGNLYGAYVPSLALSTL 331
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
D 233 RSKAARMVVAVFAVPSLPHITKTAVLAVRSTPGVSCPVLETFAAAYKGTFRPASA 292
QY 332 NSCVDFPIIYVSAERDKVRAGLFQSPGD 362
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
D 293 NSVLDPIFYFTQOKR- - - - -RQPHD 314
RESULT 15

CSAR_MACMU STANDARD; PRT; 340 AA.
ID AC P79188;
DT 15-JUL-1998 (REL. 36, CREATED)
DT 15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE C5A ANAPHYLATOXIN CHEMOTACTIC RECEPTOR (C5A-R) (FRAGMENT).
GN C5R1 OR C5AR.
OS MACACA MULATTA (RHESUS MACAQUE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC PRIMATES; CATARRHINI; CERCOPITHECIDAE; CERCOPITHECINAE; MACACA.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 96421539.
RA ALVAREZ V., COTO E., SEHEN F., GOUZALEK-KOCES S., LOPEZ-LARREA C.;
RT "Molecular evolution of the N-formyl peptide and C5a receptors in
RT non-human primates";
RL IMMUNOGENETICS 44:446-452(1996).
CC -!- FUNCTION: RECEPTOR FOR THE CHEMOTACTIC AND INFLAMMATORY PEPTIDE
CC ANAPHYLATOXIN C5A. THIS RECEPTOR STIMULATES CHEMOTAXIS, GRANULE
CC ENZYME RELEASE AND SUPEROXIDE ANION PRODUCTION.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: X97731; E242336; -
CC GCRDB: GCR_1743; -
CC PROSITE: PS00237; G_PROTEIN_RECEPTOR; 1.
CC PFAM: PF00001; 7tm_1; 1.
CC G-PROTEIN COUPLED RECEPTOR; TRANSMEMBRANE; GLYCOPROTEIN;
KW CHEMOTAXIS.
FT NON_TER 1 1
FT DOMAIN <1 30 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 31 53 1 (POTENTIAL).
FT DOMAIN 54 64 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 65 87 2 (POTENTIAL).
FT DOMAIN 88 103 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 104 125 3 (POTENTIAL).
FT DOMAIN 126 146 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 147 167 4 (POTENTIAL).
FT DOMAIN 168 193 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 194 219 5 (POTENTIAL).
FT DOMAIN 220 235 6 (POTENTIAL).
FT TRANSMEM 236 258 7 (POTENTIAL).
FT DOMAIN 259 275 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 276 296 7 (POTENTIAL).
FT DOMAIN 297 >340 CYTOPLASMIC (POTENTIAL).
FT DISULFID 102 181 BY SIMILARITY.
FT NON_TER 340 340
SQ SEQUENCE 340 AA; 38274 MW; 6FD1B699 CRC32;

Query Match 18.4%; Score 366; DB 1; Length 340;
Best Local Similarity 30.7%; Pred. No. 3.2e-17;
Matches 107; Conservative 57; Mismatches 132; Indels 52; Gaps 11;
QY 56 NDSDTLE- - - - -LPDSSRALLLGWVPTRLVPALYGLVVLVGLPANGALWVLATQAPRLPS 111
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
D 8 DDKDITLDANTPVDKTSNTLR- - -VPDILALVIFAVVFLVGLRNLVWVWTFEAKRTIN 64
QY 112 TMLNMLNATADLLALALPPRIAYHLRGQRPFGEEAARLATAALYGHMYSVLLAAVS 171
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
D 65 AIWFLNLAVALDFLSCLPLILETSIVQHHHWPFGGAACRLPSLLNLLNMYASILLATIS 124
QY 172 LDRYLAHVPLRARALRGRRRLALGLCMAWLMMAALALPLTLQRTFLARSD- - - - -RVL 227
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
RESULT 15

Db 125 ADRFLLVFNPIWQNFGRGAGLAWIACAVAW-----GLALLLTIPSLYRVVREYFPPKVL 180
Qy 228 C-----HDALPLDAQASHWQPAFTCLALLGCFPLLAMLLCYG-ATLHTLAASGRRYGHA 281
Db 181 CGVDHGHDKRRERAVA-----IARLVLGFWYPLTLTMCYTFLLRLTWSRRATRSTKT 233
Qy 282 LRLTAVVLASAVAFVPSNLL-LLLHSDPSPSANGNLYGAVVPSIALS--TLNSCVDPF 338
Db 234 LKVVAVVASFFIFWLPYQVTGMMMSFLEPSSPTF--LLKKLDSLCSIFAYINCCINPI 291
Qy 339 IYYVSAEERDKVRAGL-----FORSPGDTVASK 367
Db 292 IYVAGQFGQRLRKSLPSSLRLNVLTEESMVRESKSFTRSTVDVTMAQK 339

Search completed: October 29, 1999, 15:52:27
Job time: 17104 sec


```

Db 72 TGKLTTFVPIVYIVGILPNSGMALWIFLFTRTKKKKHPAVIYMANLALADLLSVIWF 131
Qy 131 PRIAYHLRGORWPFGEACRLATRAALGHMYGVSLLAAVSLDORYLALVHPL---RARAL 187
Db 132 LKISYHLGNWNYGEALCKVYLGIGFTYGNMFCSLFTCLSVQRYWYIVAPMGHPRKA- 191
Qy 188 RGRRLALGLCMAAWLMAAALPLTLQRTQFRLARSORVLCHDALPLDAQASHWQPAFTC 247
Db 191 ---NIAVGSALWLLIFLTVIPLVVMQTIYIPALNITTCDDVLPEEVLVGDMEYFLS 247
Qy 248 LALLGCFL-PLLAMLLCYGATLHTLAAG-----RRYGHALRLTAVVLASAVAFFVPSN 300
Db 248 LA-IGVFLFPALLTASAYVLMIKTLRSAMDHESEKRRQAIRLIITVLAMYFICEFAPS 306
Qy 301 LLLLLHYSDPSSAWGNLYGAYVPSLALSTLNSGVDPFIYVYVYSAEPRDKVRAGLFORS- 360
Db 307 LLLVVHFLIKTORQSHVYALYLALCLSLTNSCIDFVYFYVYFSKDFRDHARNALLCRSV 366
Qy 360 -PGDTVASKASAEGRSGMGTHSS 382
Db 367 RTVNRMQISLSSNFKSRKSGSYSS 390

RESULT 3
US-08-476-976-61
: Sequence 61, Application US/08476976
: Patent No. 5874400
: GENERAL INFORMATION:
: APPLICANT: SUNDELIN, JOHAN
: APPLICANT: SCARBOROUGH, ROBERT W.
: TITLE OF INVENTION: RECOMBINANT C140 RECEPTOR, ITS AGONISTS
: TITLE OF INVENTION: AND ANTAGONISTS, AND NUCLEIC ACIDS ENCODING THE RECEPTOR
: NUMBER OF SEQUENCES: 63
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: MORRISON & FOERSTER
: STREET: 2000 Pennsylvania Ave. N.W., Ste. 5500
: CITY: Washington
: STATE: D.C.
: COUNTRY: USA
: ZIP: 20006-1812
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/476,976
: FILING DATE: 07-JUN-1995
: CLASSIFICATION: 536
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/390,301
: FILING DATE: 25-JAN-1995
: ATTORNEY/AGENT INFORMATION:
: NAME: ADLER, REID G.
: REGISTRATION NUMBER: 30,988
: REFERENCE/DOCKET NUMBER: 2803-0006.20
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (202) 887-1500
: TELEFAX: (202) 887-0763
: TELEX: 90-4030
: INFORMATION FOR SEQ ID NO: 61:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 399 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
US-08-476-976-61

```

Query Match 26.6%; Score 530; DB 2; Length 399;
Best Local Similarity 34.1%; Pred. No. 1.3e-34;
Matches 131; Conservative 68; Mismatches 141; Indels 44; Gaps 11;

Matches	131;	Conservative	67;	Mismatches	142;	Indels	44;	Gaps	11;
QY	13	GFSLSGGTOT-PSVYDESGTGGDDSTPSILPAPRGYPGQVCANDSTLELPDSSRALL	71						
Db	33	GRSLIGRLTQPPI-----TGKG-----VPVEPGF-----SIDFASAIL	67						
QY	72	LGVNPTRLPALYGLVVLGVPANGALAW-LATQAPRLPSTMLLMNLATADLLALALAL	130						
Db	68	TGKLTVFLPVYIIVFVIGLPGNHALWIFLFRKKKHFAVIYMANLALADLLSVTFWP	127						
QY	131	PRIAYHLRGORWPFGEAACRLAATYALYGHMYGVSLLAAVSLDRYALVHPL---	187						
Db	128	LKTSYHLHGNNWYGERALCKVLGFFGYGMYCSILFMTCLSVQFYWYIVNPMGPRKA-	187						
QY	188	RGRRLALGLCMAAWLMAAALALPLTLQORTFLARSDRVLCSDALPLDAQASHWQPAFC	247						
Db	187	--NIAVGVSIAWLLIFLVTIPLYVMQIYIIPALNITTCDDVPLEVLGDMFNFLS	243						
QY	248	LALLGFL-PLLAMLICYGATLHTLAASG-----RRYGHALRLTAVVLASAVAFFVPSN	300						
Db	244	LA-IGVFLPALLTASAYVLMKLTSSAMDEHSENKRQAIRLIITVLAMYFICFAPS	302						
QY	301	LLLLLHSDSPSAWGNLYGAYVPSIALSTLNSCVDPIFYIYSAEPRDKVRAGLFORS-	360						
Db	303	LLLVWHYFLIKTQRSHRYALYLVALCLSTLNSCIDPFYIYFVSKDFRDHARNALLCRSV	362						
QY	360	-PGDTVASKASAEGRSGMGTHSS	382						
Db	363	RTVNRMQISLSSNNKFRKSGSYSS	386						

RESULT 5
US-08-476-000-2
: Sequence 2, Application US/08476000
: Patent No. 5716789
: GENERAL INFORMATION:
: APPLICANT: SUNDELIN, JOHAN
: APPLICANT: SCARBOROUGH, ROBERT M.
: TITLE OF INVENTION: RECOMBINANT C140 RECEPTOR, ITS AGONISTS
: AND ANTAGONISTS, AND NUCLEIC ACIDS ENCODING THE RECEPTOR
: NUMBER OF SEQUENCES: 63
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: MORRISON & FOERSTER
: STREET: 2000 Pennsylvania Ave. N.W., Ste. 5500

CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20006-1812
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/476.000
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/390.301
FILING DATE: 25-JAN-1995
ATTORNEY/AGENT INFORMATION:
NAME: ADLER, REID G.
REGISTRATION NUMBER: 30,988
REFERENCES/DOCKET NUMBER: 2803-0006.20
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 887-1500
TELEFAX: (202) 887-0763
TELEX: 90-4030
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 395 amino acids
TYPE: amino acid
TOPOLOGY: linear

Query Match 26.5%; Score 528; DB 1; Length 395;
Best Local Similarity 34.1%; Pred. No. 1.9e-34;

; MOLECULE TYPE: protein
US-08-476-000-2

Query Match 26.5%; Score 528; DB 1; Length 395;
Best Local Similarity 34.1%; Pred. NO. 1.9e-34;
Matches 131; Conservative 67; Mismatches 142; Indels 44; Gaps 11;

Qy	13	GFSLSGGTQT-PSVYDESGSTGGDDSTPSILPAPRGYPGQVCANDSDTLELPDSSRALL	71
Db	33	GRSLIGRLTQPI-----TGKG-----VPEPGF-----SIDFASAIL	67
Qy	72	LGWPTRLVALGLVLVGLPANGALW-LATQAPRLPSTMLLMNLATADLLALAP	130
Db	68	TGKLTTVFLPVYIIVFVIGLSPGNMALWIFLPTKKKHPAVIYMANLALADLSWIWPP	127
Qy	131	PRIAHYLGRGWPFGEACRLATAALYGHMYGSVLLAAVSLDRYDALVHPL---RARAL	187
Db	128	LKISYHLHGNNWYVEALCKVLIGFFGYNNYCSILFWTCLSVQRYWYVYNMGHPKKA-	187
Qy	188	RGRLALGLCMAWLMAAALALPITLQRTQFLRARDVLCHDALPDLQAASHWQPAFTC	247
Db	187	---NIAVGSALWLLFLFTIPLYNMKQTIYIPALNITTHCPDVEEVLVGMENYFLS	243
Qy	248	LALIGCFE-PLLAMLLCYGATLHTLAASG-----RRYGHALLETAVLASAVAFVPSN	300
Db	244	LA-IGVFFPALITASAYVLMITLHSSAMDEHSENKRQRAILITVILAWYFICFAPS	302
Qy	301	LLLLLHSDPSGAWNGLYAYVPSLALSTLNSCVDPFIYIYVYSAEFRDKVRAGLFORS-	360
Db	303	LLLWVHFYLIKTQSHVYALYLVALCLSTLNSCIDPFYIYFYFKDFRHNALLCRSV	362
Qy	360	-PGDTVASKASAGGSRGMGTHSS	382
Db	363	RTVNRMQISLSSNNKFSRKSGSYSS	386

RESULT 6

US-08-472-840-2
: Sequence 2, Application US/08472840
: Patent No. 5763575
: GENERAL INFORMATION:
: APPLICANT: SUNDELIN, JOHAN
: APPLICANT: SCARBOROUGH, ROBERT M.
: TITLE OF INVENTION: RECOMBINANT C140 RECEPTOR, ITS AGONISTS
: TITLE OF INVENTION: AND ANTAGONISTS, AND NUCLEIC ACIDS ENCODING THE RECEPTOR
: NUMBER OF SEQUENCES: 63
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: MORRISON & FOERSTER
: STREET: 2000 Pennsylvania Ave. N.W., Ste. 5500
: CITY: Washington
: STATE: D.C.
: COUNTRY: USA
: ZIP: 20006-1812
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/472,840
: FILING DATE:
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US/08/390,301
: FILING DATE: 25-JAN-1995
: ATTORNEY/AGENT INFORMATION:
: NAME: ADLER, REID G.
: REGISTRATION NUMBER: 30,988
: REFERENCE/DOCKET NUMBER: 2803-0006.20
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (202) 887-1500
: TELEFAX: (202) 887-0763

```

;      TELEX: 90-4030
;      INFORMATION FOR SEQ ID NO: 2:
;      SEQUENCE CHARACTERISTICS:
;      LENGTH: 395 amino acids
;      TYPE: amino acid
;      TOPOLOGY: linear
;      MOLECULE TYPE: protein
US-08-472-840-2

```

Query Match	26.5%	Score 528;	DB 2;	Length 395;
Best Local Similarity	34.1%	Pred. No. 1.9e-34;		
Matches 131; Conservative	67;	Mismatches 142;	Indels 44;	Gaps 11;

Qy	13	GFSLSGGTQT~PSVYDESGSTGGDDSTPILPAPRGYPGVGCANDSDTLLELPDSSRALL	71
Db	33	GRSLIGRLTEOPPI-----TGKG-----VPVEPGF-----SIDESASIL	67
Qy	72	LGWPTPLRVPALXGLVLVVGLPANGALAW-LATQA RPLSTMLLMNLTADLLALLALP	130
Db	68	TGKTUTTFELPVVIIVFI GLPNSGMALWIFLFTKKKHPAVIYMANLADLLUSVTWF	127
Qy	131	PRIAYHGRQRWPGEAACRLATAALXCHMYGSVLLLAAYSLDRYLALVHL~---RARAL	18
Db	128	LKISYHLHGNNWVYGEALCKVLGFFGYGNMYSILFWCLSVQRYWYIVNPMGHPRKA-	18
Qy	188	RGRRLALGLCMAAWLMAAALPUTLQRTQTRLARSDRVLC HDALPDADAQASHWQPATC	247
Db	187	---NIAVGS LAIWLII FLVTIPLYVMKOTIYIPALNITTC HDPEEVLVDGMFNFLS	24
Qy	248	LALLGCCL-PLLAMLLCYGATLHLTAASG-----RRYGHARLRTAVVLASAVAFFVPSN	300
Db	244	LA-IGVF LFPALLTASAYVLMIKTLRSSAMDEHSENKRQRAIRLIITVLAMYFTCFAPS	300
Qy	301	LLLLHHSDPSPSAWGNIAGYVPSLALSTLNSCVDPFIYYYSAAEFKDKVRACLGFOR-	360
Db	303	LLL VVHFELIKTRQSHVALYLVALCLSTLNSCIDFVYFYFSKDFDHARNALLCRSV	36
Qy	360	-PGDTVASKASAEGSRGMGTHSS	382
Db	363	RTVNRMQIISLSSNKFRSKSGSYSS	386

RESULT

US-08-476-976-2
Sequence 2, Application US/08476976
Patent No. 587400
GENERAL INFORMATION:
APPLICANT: SUNDELIN, JOHAN
APPLICANT: SCARBOROUGH, ROBERT M.
TITLE OF INVENTION: RECOMBINANT C140 RECEPTOR, ITS AGONISTS
AND ANTAGONISTS, AND NUCLEIC ACIDS ENCODING THE RECEPTOR
NUMBER OF SEQUENCES: 63
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORRISON & FOERSTER
STREET: 2000 Pennsylvania Ave. N.W., Ste. 5500
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20006-1812
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/476,976
FILING DATE: 07-JUN-1995
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/390,301
FILING DATE: 25-JAN-1995
ATTORNEY/AGENT INFORMATION:

NAME: ADLER, REID G.
REGISTRATION NUMBER: 30,988
REFERENCE/DOCKET NUMBER: 2803-0006.20
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 887-1500
TELEFAX: (202) 887-0763
TELEX: 90-4030
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 395 amino acids
TYPE: amino acid
MOLECULE TYPE: protein
US-08-476-976-2

Query Match 26.5%; Score 528; DB 2; Length 395;
Best Local Similarity 34.1%; Pred. No. 1.9e-34;
Matches 131; Conservative 67; Mismatches 142; Indels 44; Gaps 11;
QY 13 GFSLSGGTQT-PSVYDESGTGGDDSTPSILPAPRGYPGVQVCANDSTLELPDSSRALL 71
DB 33 GRSILGRLETPPI-----TGKG-----VPVEPGF-----SIDEFSASIL 67
QY 72 LGWPTRLVPALYGLVVLVGLPANGALMW-LATQAPRLPSTMLLMNLATADLLALLALP 130
DB 68 TGKLTTFVLPVYIIVFVIGLPSNGMALWIFLFTKKKHPAVIYMANIALADLLSVIWF 127
QY 131 PRIAYHLRGQWPGEACRLATAALYAGHYGVSLLAAVSLDRYLALVHPL---RARAL 187
DB 128 LKISYHLGNWVYGEALCKVLGFFGYGNMYCSILFWTCLSVQVWVIVNPMGHPKKA- 187
QY 188 RGRRLALGLCMAAALPLTLQRTFLARSDRVLCHDALPLDQAASHWQPAFTC 247
DB 187 ---NIAVGSALWLLIPLVITPLVYMKQTIYIPALNITTTCHDVLPEVLVGMFNFELS 243
QY 248 LALLGCEL-PLLAMLLCYGATLHTLAAG-----RRYGHALRLTAVVLASAVAFFVPSN 300
DB 244 LA-IGVFLFPALLTASAVYLMIKTLRSSAMDEHSENKQRAIRLIITVLMYFICFAPS 302
QY 301 LLLHLVSDPSAWNGLYGAVPSLALSTLNSCVDPIFYIYVSAEFRDKVRAGLFORS- 360
DB 303 LLLVHYFLIKTORSHVYALYVALCLSTLNSCIDPFIYFVSKDFRDHARNALLCSV 362
QY 360 -PGDTVASKAAGEGSRGMGTHSS 382
DB 363 RTVNRMQISLSSNFKSRKSGSYSS 386

RESULT 8
US-08-097-938-5
Sequence 5, Application US/08097938
Patent No. 5629174
GENERAL INFORMATION:
APPLICANT: SUNDELIN, JOHAN
APPLICANT: SCARBOROUGH, ROBERT M.
TITLE OF INVENTION: RECOMBINANT C140 RECEPTOR AND ITS
TITLE OF INVENTION: AGONISTS AND ANTAGONISTS
NUMBER OF SEQUENCES: 59
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORRISON & FOERSTER
STREET: 2000 Pennsylvania Ave. N.W., Ste. 5500
CITY: Washington, D.C.
COUNTRY: USA
ZIP: 20006-1812
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/097,938
FILING DATE: 26-JUL-1993

CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: MURASHIGE, KATE H.
REGISTRATION NUMBER: 29,959
REFERENCE/DOCKET NUMBER: 22803-20006.00
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 887-1500
TELEFAX: (202) 887-0763
TELEX: 90-4030
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 395 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-097-938-5

Query Match 26.4%; Score 525; DB 1; Length 395;
Best Local Similarity 34.8%; Pred. No. 3.2e-34;
Matches 125; Conservative 63; Mismatches 129; Indels 42; Gaps 10;
QY 13 GFSLSGGTQT-PSVYDESGTGGDDSTPSILPAPRGYPGVQVCANDSTLELPDSSRALL 71
DB 33 GRSILGRLETPPI-----TGKG-----VPVEPGF-----SIDEFSASIL 67
QY 72 LGWPTRLVPALYGLVVLVGLPANGALMW-LATQAPRLPSTMLLMNLATADLLALLALP 130
DB 68 TGKLTTFVLPVYIIVFVIGLPSNGMALWIFLFTKKKHPAVIYMANIALADLLSVIWF 127
QY 131 PRIAYHLRGQWPGEACRLATAALYAGHYGVSLLAAVSLDRYLALVHPL---RARAL 187
DB 128 LKISYHLGNWVYGEALCKVLGFFGYGNMYCSILFWTCLSVQVWVIVNPMGHPKKA- 187
QY 188 RGRRLALGLCMAAALPLTLQRTFLARSDRVLCHDALPLDQAASHWQPAFTC 247
DB 187 ---NIAVGSALWLLIPLVITPLVYMKQTIYIPALNITTTCHDVLPEVLVGMFNFELS 243
QY 248 LALLGCEL-PLLAMLLCYGATLHTLAAG-----RRYGHALRLTAVVLASAVAFFVPSN 300
DB 244 LA-IGVFLFPALLTASAVYLMIKTLRSSAMDEHSENKQRAIRLIITVLMYFICFAPS 302
QY 301 LLLHLVSDPSAWNGLYGAVPSLALSTLNSCVDPIFYIYVSAEFRDKVRAGLFORS 359
DB 303 LLLVHYFLIKTORSHVYALYVALCLSTLNSCIDPFIYFVSKDFRDHARNALLCS 361

RESULT 9
US-08-476-000-5
Sequence 5, Application US/08476000
Patent No. 5716789
GENERAL INFORMATION:
APPLICANT: SUNDELIN, JOHAN
APPLICANT: SCARBOROUGH, ROBERT M.
TITLE OF INVENTION: RECOMBINANT C140 RECEPTOR, ITS AGONISTS
TITLE OF INVENTION: AND ANTAGONISTS, AND NUCLEIC ACIDS ENCODING THE RECEPTOR
NUMBER OF SEQUENCES: 63
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORRISON & FOERSTER
STREET: 2000 Pennsylvania Ave. N.W., Ste. 5500
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20006-1812
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/476,000
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435

```
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/390,301
; FILING DATE: 25-JAN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: ADLER, REID G.
; REGISTRATION NUMBER: 30,988
; REFERENCE/DOCKET NUMBER: 2803-0006.20
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 887-1500
; TELEFAX: (202) 887-0763
; TELEX: 90-4030
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 395 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
US-08-472-840-5

Query Match 26.4%; Score 525; DB 1; Length 395;
Best Local Similarity 34.8%; Pred. No. 3.2e-34;
Matches 125; Conservative 63; Mismatches 129; Indels 42; Gaps 10;

QY 13 GFSLSGGTQT-PSVYDESGTGGDDSTPSILPAPRGYQGVCAANDSDTLELPDSSRALL 71
Db 33 GRSLGRLETPPI-----TGKG-----VPVEPGF-----SIDEFSASIL 67
QY 72 LGWVPRLVPALYGLVLYVGLPANGIALWV-LATQAPRLPSTMLMNLATADLLALALP 130
Db 68 TGLKTLVLPVYIIIVFVIGLPSNGMALWIFLRTKKKHPAVIYMANLADLLSVINFP 127
QY 131 PRIAYHLRGQRPFGAARLATAALYGHMYGVSLLAAVSLDRYLALVHPL---RARAL 187
Db 128 LKISYHLGNWVYGCALCKVLGIFGYGNMYCSILFTCLSVQRYVWVYVNPNGHPRKKA- 187
QY 188 RGRRLALGCMALWMAALALPLTLQRTFLARSDRVLCHDALPLDAQASHWQPAFTC 247
Db 187 ---NIAVGSALWILFLVTPLYVMKQTIYPALNITTCDDVLPVEVLVGMENYFLS 243
QY 248 LALGCGFL-PLLAMLCYCATLHTLAASG-----RRYGHALRLTAVLASAVAFVPSN 300
Db 244 LA-IGVFLPALLTASAYVLMIKTLRSSAMDEHSEKKRQRAIRLIITVLAMYFICFAPS 302
QY 301 LLLLHYSDPSAWGNLYGAYVPSLALSTLNSCDVPFIYVYSAEFRDKVRAGLFQRS 359
Db 303 LLLVHYFLIKTQRQSHVYALYLVALCLSTLNSCIDPFYVYFVSKDFRDHARNALLCRS 361

RESULT 10
US-08-472-840-5
; Sequence 5, Application US/08472840
; Patent No. 5763575
; GENERAL INFORMATION:
; APPLICANT: SUNDELIN, JOHAN
; APPLICANT: SCARBOROUGH, ROBERT M.
; TITLE OF INVENTION: RECOMBINANT C140 RECEPTOR, ITS AGONISTS
; TITLE OF INVENTION: AND ANTAGONISTS, AND NUCLEIC ACIDS ENCODING THE RECEPTOR
; NUMBER OF SEQUENCES: 63
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 2000 Pennsylvania Ave. N.W., Ste. 5500
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20006-1812
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/472,840
```

```
;
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/390,301
; FILING DATE: 25-JAN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: ADLER, REID G.
; REGISTRATION NUMBER: 30,988
; REFERENCE/DOCKET NUMBER: 2803-0006.20
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 887-1500
; TELEFAX: (202) 887-0763
; TELEX: 90-4030
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 395 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
US-08-472-840-5

Query Match 26.4%; Score 525; DB 2; Length 395;
Best Local Similarity 34.8%; Pred. No. 3.2e-34;
Matches 125; Conservative 63; Mismatches 129; Indels 42; Gaps 10;

QY 13 GFSLSGGTQT-PSVYDESGTGGDDSTPSILPAPRGYQGVCAANDSDTLELPDSSRALL 71
Db 33 GRSLGRLETPPI-----TGKG-----VPVEPGF-----SIDEFSASIL 67
QY 72 LGWVPRLVPALYGLVLYVGLPANGIALWV-LATQAPRLPSTMLMNLATADLLALALP 130
Db 68 TGLKTLVLPVYIIIVFVIGLPSNGMALWIFLRTKKKHPAVIYMANLADLLSVINFP 127
QY 131 PRIAYHLRGQRPFGAARLATAALYGHMYGVSLLAAVSLDRYLALVHPL---RARAL 187
Db 128 LKISYHLGNWVYGCALCKVLGIFGYGNMYCSILFTCLSVQRYVWVYVNPNGHPRKKA- 187
QY 188 RGRRLALGCMALWMAALALPLTLQRTFLARSDRVLCHDALPLDAQASHWQPAFTC 247
Db 187 ---NIAVGSALWILFLVTPLYVMKQTIYPALNITTCDDVLPVEVLVGMENYFLS 243
QY 248 LALGCGFL-PLLAMLCYCATLHTLAASG-----RRYGHALRLTAVLASAVAFVPSN 300
Db 244 LA-IGVFLPALLTASAYVLMIKTLRSSAMDEHSEKKRQRAIRLIITVLAMYFICFAPS 302
QY 301 LLLLHYSDPSAWGNLYGAYVPSLALSTLNSCDVPFIYVYSAEFRDKVRAGLFQRS 359
Db 303 LLLVHYFLIKTQRQSHVYALYLVALCLSTLNSCIDPFYVYFVSKDFRDHARNALLCRS 361

RESULT 11
US-08-472-840-5
; Sequence 5, Application US/08476976
; Patent No. 5874400
; GENERAL INFORMATION:
; APPLICANT: SUNDELIN, JOHAN
; APPLICANT: SCARBOROUGH, ROBERT M.
; TITLE OF INVENTION: RECOMBINANT C140 RECEPTOR, ITS AGONISTS
; TITLE OF INVENTION: AND ANTAGONISTS, AND NUCLEIC ACIDS ENCODING THE RECEPTOR
; NUMBER OF SEQUENCES: 63
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 2000 Pennsylvania Ave. N.W., Ste. 5500
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20006-1812
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
```


;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/476,976
;; FILING DATE: 07-JUN-1995
;; CLASSIFICATION: 536
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 08/390,301
;; FILING DATE: 25-JAN-1995
;; ATTORNEY/AGENT INFORMATION:
;; NAME: ADLER, REID G.
;; REGISTRATION NUMBER: 30,988
;; REFERENCE/DOCKET NUMBER: 2803-0006.20
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (202) 887-1500
;; TELEFAX: (202) 887-0763
;; TELEX: 90-4030
;; INFORMATION FOR SEQ ID NO: 5:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 395 amino acids
;; TYPE: amino acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; US-08-476-976-5

Query Match 26.4%; Score 525; DB 2; Length 395;
Best Local Similarity 34.8%; Pred. No. 3.2e-34;
Matches 125; Conservative 63; Mismatches 129; Indels 42; Gaps 10;

QY 13 GFSLSGGTQ-PSYVDESGTGGDDSTPSILPAPRGVPGVCANDSTLELPDSSRALL 71
Db 33 GRSILGLETOPPI-----TGKG-----VPVEPGF-----SIDFSAIIL 67
QY 72 LGWVPTRLVPALYGLVYVGLPANGALWV-LATQAPRLPSTMLMNLATADLLALALP 130
Db 68 TGKLTTFVLPVYIIVFVIGLPSMGALWIEFTKTKKHPAIVYMANLALDLSLWIFP 127
QY 131 PRIAYHLRGQWPGEACRATATAALYGHMYGVSLLLAASLDYRLALVHPL---RARAL 187
Db 128 LKISYHLGNWVYGEALCKVLIGFFYGNMYCSILFTCLSVQRYWIVNPMGHPKKA- 187
QY 188 RGRRLALGLCHMAAALALPALTQRTPLARSQVLCVLDALPLDAQASHWQPAFTC 247
Db 187 ---NIAGVSLAIWLLFLVPIYVVKQTIYIPALNITTCVLDVPEVLVGMFNFELS 243
QY 248 LALLGCEL-PLLAMLLCYGATLHTLAASG-----RRYGHALRLTAVVLASAVAFFVPSN 300
Db 244 LA-IGVFLPALLTASAVLMIKTLRSSAMDEHSEKKRQRAIRLIITVLMYFICFAPS 302
QY 301 LLLLLHYSDPSPSANGNLYGAVVPSLALSTLNSCVDPFIYVYVNAEPDKYRACLFOES 359
Db 303 LLLVVHYFLIKTRQSHVYALYVALCLSTLNSCIDPFYVYVSKDFRDHARNALLCRS 361

RESULT 12
US-08-742-440A-6
; Sequence 6, Application US/08742440A
; Patent No. 5892014
; GENERAL INFORMATION:
; APPLICANT: Coughlin, Shaun
; APPLICANT: Ishihari, Hiroaki
; APPLICANT: Connolly, Andrew
; TITLE OF INVENTION: Protease Activated Receptor
; TITLE OF INVENTION: 3 and Uses Thereof
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Bozicevic & Reed, LLP
; STREET: 285 Hamilton Avenue, Suite 200
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94301
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette

;; COMPUTER: IBM Compatible
;; OPERATING SYSTEM: DOS
;; SOFTWARE: FastSeq for Windows Version 2.0
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/742,440A
;; FILING DATE: 30-OCT-1996
;; CLASSIFICATION: 536
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER:
;; FILING DATE:
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Sherwood, Pamela J
;; REGISTRATION NUMBER: 36,677
;; REFERENCE/DOCKET NUMBER: UCAL/060PAT
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 650-327-3400
;; TELEFAX: 650 327-3231
;; TELEX:
;; INFORMATION FOR SEQ ID NO: 6:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 408 amino acids
;; TYPE: amino acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; FEATURE:
;; US-08-742-440A-6

Query Match 26.2%; Score 521.5; DB 2; Length 408;

Best Local Similarity 33.8%; Pred. No. 6.2e-34;
Matches 120; Conservative 62; Mismatches 158; Indels 15; Gaps 5;
QY 3 GRLLMLVLVGLFSLGGTQ-----TPSYVDESGTGGDDSTPSILPAPRGVQV---- 54
Db 29 GLULLLTFQCSGMENDNNLAKPTLPKTFRGAPPNSFEFP--FSALEGWGTATIVK 86
QY 54 --CANDSDT--LELPDSSRALLGWVPTRLVPALYGLVYVGLPANGALWVLTQAPRLP 110
Db 87 IKCPESASLHVKNATMGYLTSLSTKLIPAIYLLVYVGVPAVYVTLMLFFERTSIC 146
QY 111 STMLMLNATADLLALPRLIAYHLRGQWPGEACRATATAALYGHMYGVSLLLAASV 170
Db 147 TTYFTYNLATADLFECVTLFPKTAHNGNWNWVGEVLCRATTVIFGNYCSTILLACI 206
QY 171 SLDRYALVHPLRARALRGRLALGLCMAAALALPLTLQRTPLARSQVLCVLD 230
Db 207 SINRYLAIVHPFTYRGVLPKHTYALVTCGLVWATVFLYMLPFFILKQEVYLVQPDITTC 266
QY 231 ALPLDAQASHWQ-PAFTCLALLGCFPLLAMLLCYGATLHTLAASGRYGHALRLTAVVL 289
Db 267 VHNTESSSPFQLYYFISLAFPGFLIPFIYIYCYAAIIRTLNAYDHRWLMYVKASLLIL 326
QY 290 ASAVAFVPSNLLLLHYSDPSPSANGNLYGAVVPSLALSTLNSCVDPFIYVYV 344
Db 327 VIPTICFAPSNIILIIHHANYYNNTDGLFIYIILALCLGSLNSCLDPFFLYFLMS 381

RESULT 13
US-08-097-938-6
; Sequence 6, Application US/08097938
; Patent No. 5623174
; GENERAL INFORMATION:
; APPLICANT: SUNDELIN, JOHAN
; APPLICANT: SCARBOROUGH, ROBERT M.
; TITLE OF INVENTION: RECOMBINANT C140 RECEPTOR AND ITS
; TITLE OF INVENTION: AGONISTS AND ANTAGONISTS
; NUMBER OF SEQUENCES: 59
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 2000 Pennsylvania Ave. N.W., Ste. 5500
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20006-1812

```

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/097,938
; FILING DATE: 26-JUL-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: MURASHIGE, KATE H.
; REGISTRATION NUMBER: 29,959
; REFERENCE/DOCKET NUMBER: 22803-20006.00
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 887-1500
; TELEFAX: (202) 887-0763
; TELEX: 90-4030
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 398 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-097-938-6

```

Query Match	25.6%	Score	510;	DB	1;	Length	398;
Best Local Similarity	36.5%;	Pred.	No. 4.9e-33;				
Matches	110;	Conservative	58;	Mismatches	123;	Indels	10;
Gaps	5;						

QY	67	SRALLCGVTRLPALYGLVVLVGLPANGALWV-LATQAPRLPSTMLNMLNATADLLL	125
DB	66	SASVLTGKLTVPUPYIIVVVGLPNGMALWFLFTRKKHPAVYIMANLALADLLS	125
QY	126	ALALPPRIAYHLRGORPFGGEACRLATAALYGHMYGSVLLLAASVLDLYALVHPLRAR	185
DB	126	VINEPLKIAYHIHGNMYIGBALCNVLIGFFYGNMYCSILEFTCLSVORYVYVNP-M-GH	184
QY	186	ALRGRRALGCMAAWLMAALPLTLQROTFLRSLDRVLCHDALPDAQASHWQPAF	245
DB	185	SRKANIAIGISLAIMULLLLVTPLYVVKOTIFIPALNITTHCDVLPQEQVLGDMFNYF	244
QY	246	TCIALLCGFL-PLLAMLLCYGATLHTLAASG-----RRYGHALRELTAVVLASAVAFVP	298
DB	245	LSLA-IGVFLFPFALTASAVYLMIRMRSSAMDENSEKKRKRAIKLIVTVLAWLYCIFP	303
QY	299	SNLLLLHYDPSWAGNLYGAVPSLALSTLNSCDVPFIYVYVSAREFRDKVRAGLQF	358
DB	304	SNLLVVHYFLIKSQGSHVYALIVALCLSTLNSCIDPFYVYVSHDFRDHAKNALLCR	363
QY	359	S	359
DB	364	S	364

```

RESULT 14
US-08-476-000-6
; Sequence 6, Application US/08476000
; Patent No. 5716789
; GENERAL INFORMATION:
; APPLICANT: SUNDELIN, JOHAN
; APPLICANT: SCARBOROUGH, ROBERT M.
; TITLE OF INVENTION: RECOMBINANT C140 RECEPTOR, ITS AGONISTS
; TITLE OF INVENTION: AND ANTAGONISTS, AND NUCLEIC ACIDS ENCODING THE RECEPTOR
; NUMBER OF SEQUENCES: 63
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 2000 Pennsylvania Ave. N.W., Ste. 5500
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20006-1812
; COMPUTER READABLE FORM:

```

```

; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/476,000
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/390,301
; FILING DATE: 25-JAN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: ADLER, REID G.
; REGISTRATION NUMBER: 30,988
; REFERENCE/DOCKET NUMBER: 2803-0006.20
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 887-1500
; TELEFAX: (202) 887-0763
; TELEX: 90-4030
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 398 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-476-000-6

Query Match 25.6%; Score 510; DB 1; Length 398;
Best Local Similarity 36.5%; Pred. No. 4.9e-33;
Matches 110; Conservative 58; Mismatches 123; Indels 10; Gaps 5;

Qy 67 SRALLGWVTRLPALYGLVYVGLPANGALWV-LATQAPRLPSTMLLNLTADLL 125
Db |::||| |::||| |::||| |::||| |::||| |::||| |::||| |::|||
Qy 126 ALALPPRIAYHLRQRPFFGEACRLATAALYGHMYGSLVLLAAVSLDRYLALVHPLR 185
Db |::||| |::||| |::||| |::||| |::||| |::||| |::||| |::|||
Qy 126 VWFPLKIAYHIHGNWNYGEALCNVLGFFYGNMYCSILFMTCLSVORYWVIVNPM-GH 184
Db |::||| |::||| |::||| |::||| |::||| |::||| |::||| |::|||
Qy 186 ALRGERRALGICMAWLAALPLTLQRTFLARSDRVLCHDALPLDAQASHWQPAF 245
Db |::||| |::||| |::||| |::||| |::||| |::||| |::||| |::|||
Qy 185 SRKKAITAIGSLAIWLLITVPLVYVVKQTIFIPALNITTCHDVLPQVLYGDMFN 244
Db |::||| |::||| |::||| |::||| |::||| |::||| |::||| |::|||
Qy 246 TCLALLGCLF-LILAMLLCYGATLHTLAASG-----RRYGHALRLTAVLASAVAF 298
Db ||::||| |::||| |::||| |::||| |::||| |::||| |::||| |::|||
Qy 245 LSLA-IGVLEPAPLTASAVYLMRLRSSAMDENSEKKRKRAIKLIVTAWLYLICFIP 303
Db |::||| |::||| |::||| |::||| |::||| |::||| |::||| |::|||
Qy 299 SNLLLLHYDSPSANGNYGAYVPSLALSTLNSCVDDPFIIYVVSAAEFKDKVRAGLF 358
Db |::||| |::||| |::||| |::||| |::||| |::||| |::||| |::|||
Qy 304 SNLLLVHYFLIKSQGSHVYALYIVALCLSTLNSCIDPFYFVYVSHDFRDKARNALL 363
Db |::||| |::||| |::||| |::||| |::||| |::||| |::||| |::|||
Qy 359 S 359
Db 364 S 364

RESULT 15
US-08-472-840-6
; Sequence 6, Application US/08472840
; Patent No. 5763575
; GENERAL INFORMATION:
; APPLICANT: SUNDELIN, JOHAN
; APPLICANT: SCARBOROUGH, ROBERT M.
; TITLE OF INVENTION: RECOMBINANT C140 RECEPTOR, ITS AGONISTS
; TITLE OF INVENTION: AND ANTAGONISTS, AND NUCLEIC ACIDS ENCODING THE RECEPTOR
; NUMBER OF SEQUENCES: 63
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 2000 Pennsylvania Ave. N.W., Ste. 5500
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA

```

ZIP: 20006-1812
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/472,840
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/390,301
FILING DATE: 25-JAN-1995
ATTORNEY/AGENT INFORMATION:
NAME: ADLER, REID G.
REGISTRATION NUMBER: 30,988
REFERENCE/DOCKET NUMBER: 2803-0006.20
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 887-1500
TELEFAX: (202) 887-0763
TELEX: 90-4030
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 398 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-472-840-6

Query Match 25.6%; Score 510; DB 2; Length 398;
Best Local Similarity 36.5%; Pred. No. 4.9e-33;
Matches 110; Conservative 58; Mismatches 123; Indels 10; Gaps 5;
QY 67 SRALLGWPTRLVPLVGVVGLPANGIALV-LATQAPRLPSTMLMLNLTADLL 125
DB 66 SASVLTKLTVPVPIVYIIVFVGLPSGMLVFLRTKKKHPAVIYMANLADLLS 125
QY 126 ALALPPRIAYHLRGQRPFGAEACRLATAALYGHMYGSVLLLAASVLDYALVHPLAR 185
DB 126 VWFPLKIAVHIGNNWYGEALCNVLIGFFYGNMYCSILFTCLSVQRYWVIVNPW-GH 184
QY 186 ALGRRLALGLCAAWLMAALPLTLQRTFLARSDRVLCCHDALPLDAQASHWQPAF 245
DB 185 SRKANAIAGISLAIWLLILLVPLVVKQTIIPALNITTCDDVLPQVLYGDMENYF 244
QY 246 TCLALLGCFL-PLIALLCYCATLHTLAAG-----RRYCHALRLTAVVLASAVAFVFP 298
DB 245 LSLA-IGVFLPFLTASAVLMIRLSSAMDENSEKKRKAIRKLIYTVLAMYLCIFP 303
QY 299 SNLLLLHYSDPSPSAGNLYGAYVPSLALSTLNSCVDPFIYYVSAEFRDKVRAGLFQR 358
DB 304 SNLLVHYFLIKSQGSHVYALYVALCLSTLNSCIDPEVYFVSHDFRDHAKNALLCR 363
QY 359 S 359
DB 364 S 364

Search completed: October 31, 1999, 06:15:14
Job time: 2194 sec

This Page Blank (uspio)

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 31, 1999, 00:52:02 ; Search time 73.12 Seconds
(without alignments)
6965.158 Million cell updates/sec

Title: US-09-371-333-1
Perfect score: 4895
Sequence: 1 CTCACCGGCTGTGGCA.....AAAAAAAAAAAAAAAAAAAA 4895

Scoring table: IDENTITY_NUC

Searched: 192659 seqs, 52021692 residues

Database : Issued Patents_NA.*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/1/ina/5C_COMB.seq.*
4: /cgn2_6/ptodata/1/ina/5D_COMB.seq.*
5: /cgn2_6/ptodata/1/ina/PTUS9_COMB.seq.*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	381.2	7.8	31571	1 US-08-323-443B-1	Sequence 1, Appl
C 2	336.4	6.9	8342	5 PCT-US94-04496-63	Sequence 63, Appl
C 3	336.2	6.9	11613	1 US-08-484-044-10	Sequence 10, Appl
C 4	324.6	6.6	8392	1 US-08-080-255-6	Sequence 6, Appl
C 5	324.6	6.6	8392	5 PCT-US93-05857-6	Sequence 6, Appl
C 6	314.4	6.4	22481	5 PCT-US95-07201-43	Sequence 43, Appl
C 7	306.6	6.3	8174	1 US-07-914-281-5	Sequence 5, Appl
C 8	306.6	6.3	8174	1 US-08-393-246-5	Sequence 5, Appl
C 9	306.6	6.3	3373	1 US-08-273-411-2	Sequence 2, Appl
C 10	306.6	6.3	8174	3 US-08-525-058A-5	Sequence 5, Appl
C 11	306.6	6.3	8174	4 US-08-696-731-5	Sequence 5, Appl
C 12	306.6	6.3	8174	5 PCT-US91-00899-3	Sequence 3, Appl
C 13	299.6	6.1	4421	3 US-08-257-963B-9	Sequence 9, Appl
C 14	299.6	6.1	4421	5 PCT-US95-07201-9	Sequence 9, Appl
C 15	296.6	6.1	2923	2 US-08-480-449-1	Sequence 1, Appl
C 16	296.6	6.1	2923	4 US-08-660-542-1	Sequence 1, Appl
C 17	295.6	6.0	6769	2 US-08-480-784-20	Sequence 20, Appl
C 18	295.6	6.0	6769	2 US-08-483-553-20	Sequence 20, Appl
C 19	295.6	6.0	6769	2 US-08-487-002-20	Sequence 20, Appl
C 20	295.6	6.0	6769	2 US-08-483-554B-20	Sequence 20, Appl
C 21	295.6	6.0	6769	2 US-08-488-011B-20	Sequence 20, Appl
C 22	295.6	6.0	6769	5 PCT-US95-10202-20	Sequence 20, Appl
C 23	295.6	6.0	6769	5 PCT-US95-10203-20	Sequence 20, Appl
C 24	295.6	6.0	6769	5 PCT-US95-10220-20	Sequence 20, Appl
C 25	288	5.9	246240	4 US-08-724-394A-20	Sequence 20, Appl
C 26	288	5.9	246240	4 US-08-724-394A-21	Sequence 21, Appl
C 27	288	5.9	246240	4 US-08-724-394A-22	Sequence 22, Appl
C 28	281	5.7	7620	1 US-07-767-135-1	Sequence 1, Appl
C 29	281	5.7	7620	1 US-07-841-652-1	Sequence 1, Appl
C 30	279.6	5.7	7676	3 US-08-451-777A-7	Sequence 7, Appl
C 31	279.6	5.7	7676	3 US-08-451-778A-7	Sequence 7, Appl
C 32	279.6	5.7	7676	5 US-08-998-208-7	Sequence 7, Appl
C 33	279.6	5.7	7676	5 PCT-US95-06743-7	Sequence 7, Appl
C 34	278.8	5.7	1442	3 US-08-454-557C-120	Sequence 120, App
C 35	278.8	5.7	1442	4 US-08-340-426D-120	Sequence 120, App
C 36	278.8	5.7	1442	4 US-08-450-673C-120	Sequence 120, App
C 37	277.6	5.7	4421	3 US-08-257-963B-9	Sequence 9, Appl

38 277.6 5.7 4421 5 PCT-US95-07201-9 Sequence 9, Appl
39 275.8 5.6 31571 1 US-08-323-443B-1 Sequence 1, Appl
40 270.8 5.5 2713 4 US-08-916-901-6 Sequence 6, Appl
C 41 269.4 5.5 3035 3 US-08-726-725-2 Sequence 2, Appl
C 42 268.8 5.5 4576 3 US-08-832-883-49 Sequence 49, Appl
C 43 268.8 5.5 4576 3 US-08-832-883-49 Sequence 49, Appl
C 44 263.2 5.4 17327 1 US-07-906-871-15 Sequence 15, Appl
C 45 261.8 5.3 1418 5 PCT-US95-1711A-120 Sequence 120, App

ALIGNMENTS

RESULT 1
US-08-323-443B-1/c
; Sequence 1, Application US/08323443B
; Patent No. 5654170
; GENERAL INFORMATION:
; APPLICANT: KLINGER, KATHERINE W.
; APPLICANT: LANDES, GREGORY M.
; APPLICANT: BURN, TIMOTHY C.
; APPLICANT: CONNORS, TIMOTHY D.
; APPLICANT: DACKOWSKI, WILLIAM R.
; APPLICANT: GERMINO, GREGORY
; APPLICANT: QIAN, FENG
; TITLE OF INVENTION: POLYCYSTIC KIDNEY DISEASE GENE
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Darby & Darby PC
; STREET: 805 Third Avenue
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/323,443B
; FILING DATE: 12-OCT-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Ludwig, S. Peter
; REGISTRATION NUMBER: 25,351
; REFERENCE/DOCKET NUMBER: 0372/0A462
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 527-7700
; TELEFAX: (212) 753-6237
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 31571 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; IMMEDIATE SOURCE:
; CLONE: PKD1 GENOMIC
; US-08-323-443B-1

Query Match 7.8% Score 381.2; DB 1; Length 31571;
Best Local Similarity 67.3%; Pred. No. 1.4e-60;
Matches 719; Conservative 0; Mismatches 228; Indels 121; Gaps 8;
Qy 1534 CTCACGCGCTGTATCCCGACACTTAAGCGCCAAAGCGGATGGATGATGAGCCCGAG 1593
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 6802 CTCACACCTGTATCCCGACACTTTGGGAGCGCCGAGCGACATCACCTAAGGCCAAG 6743

Db 6757 AGGAGTTCGAGACCGAGCTGACCAAGCTGGAGAAACCTCATCTACTAAAAACAA--- 6814
Qy 1980 TTTTAATGAACAGGATTTGGCATGCGCTATAGTATCCAGCCTCAAGAGGCACAGG 2039
Db 6814 ----AATTAGCAGGATGGTGCCATATACCTGTATCCAGCTACTCAGGAGGCTGAGG 6869
Qy 2040 CGGGAGGATCACTTGAGCCTGGGAGGTTGGTTGCAGTGAGCTATGATTGTACCACTGC 2099
Db 6870 CAGGAGAATCA---GAACCCAGGAGGGAGGTTGGTTAGCTGAGATCGTGCCATTGC 6926
Qy 2100 ACTCCAGCCTGGGCAACAGCAGACAGCTTGTCTCAAAAATTAACAAACTAAAAATAAA 2159
Db 6927 ATTCCAGCCTGGGCAACAGAGTGAACCTTCATCTCAAAAAAATAAA 6978
Qy 2160 AAAGAAGACGAGAGATGAGTGGGTGGTGGCTCACACCTGCAATCCAGCAGCTTTGGAG 2219
Db 6979 AGTACTATTATACAGGCTGGGATGCTGCTACAGCTTGGAAATCCAGCAGCTTTGGGAG 7038
Qy 2220 GCGGAGGTGGGAGATCATCTGAGGCCAGGAGTTCAAGACCAGCCTGGCTAAACATGGTGA 2279
Db 7039 GCTGAAGTGGAGGATGCTTCAGCCAGGAGTTCAAGACCAGCCTGAGCAACATAATGA 7098
Qy 2280 AATCTATCTCTACCAAAAATAACAAAAT---AGCCAGGCGTGGTGGGACCTGTATC 2337
Db 7099 GACCCTGTCTACAAAAAATTAAGAAAAATCGTCCAGGAGCTGGTGGTCTGTGCCTGCAG 7158
Qy 2338 TGGGAGGTGCCACCGAGCTACTGGGAGGCTGAGTCAAGGAGATCGCTTGAACCTGG 2396
Db 7159 T-----CCTAGCTACTAGGAGTCTGAAGTAGGAGATCACTTGGAGCTGG 7204

RESULT 4

US-08-080-255-6
: Sequence 6, Application US/08080255
: Patent No. 5487970
: GENERAL INFORMATION:
: APPLICANT: Rowley, Janet D.
: APPLICANT: Diaz, Manuel O.
: TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
: TITLE OF INVENTION: DEPOSITING GENE REARRANGEMENTS AND
: TITLE OF INVENTION: TRANSLOCATIONS
: NUMBER OF SEQUENCES: 8
: CORRESPONDENCE ADDRESS:
: ADDRESS: Arnold, White & Durkee
: STREET: P. O. Box 4433
: CITY: Houston
: STATE: Texas
: COUNTRY: USA
: ZIP: 77210
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/080,255
: FILING DATE: 19930617
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Parker, David L.
: REGISTRATION NUMBER: 32,165
: REFERENCE/DOCKET NUMBER: ARCD:072/PAR
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (512) 320-7200
: TELEFAX: (512) 474-7577
: INFORMATION FOR SEQ ID NO: 6:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 8392 base pairs
: TYPE: NUCLEIC ACID
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
US-08-080-255-6

Query Match 6.58; Score 324.6; DB 1; Length 8392;
Best Local Similarity 66.9%; Pred. No. 1.6e-50;
Matches 658; Conservative 0; Mismatches 229; Indels 96; Gaps 10;

Qy 1520 GCCAGGCGCTGGTGCCTCACGCTGTAAATCCAGCAGCTTTAAGAGGCGCAAGCGGATGGAT 1579
Db 860 GCTGGGCACGGTGGCTCACGCTGGTAAATCCCAACACTTAGTGAGGCTGAGGTGGGAGAT 919
Qy 1580 CACTTGAAGCCCGAGGTTCAACACCGAGCCTGAGCAACATGTTAAACCCCTCTCTACCA 1639
Db 920 TGCCTTGAAGCCCGAGGTTCAAGACCGAGCCTGGGCAACATAGCAAGACCCCTCTCTTTATT 979
Qy 1640 AAAATACAAAAA-----TTAGCTGGGCTTGGTGGCTGGCCGCTGTATA 1680
Db 980 AAACAAAAAATAAAAGAAAGAAAGATTTAGCCAGGATGGTGGCAGTTGCGTGTATA 1039
Qy 1681 ATCCAGCTACTCAGGAGCTGAGGCGAGAGGATCG-CTTGAACCTGGGAGGCGAGAGGTT 1739
Db 1040 GTCCAGCTACTCAGGAGCTGAGATAGAGGATTTCTTGAGCCAGGAATTCAGGCT 1099
Qy 1740 GCAGTGACCCGAGATTTGGCCACCTGGAGCTCCAGCCTGGGTGACAGAGAGCCTGTCTCTAA 1799
Db 1100 GTAGTGAGCTATGATTGTACCACTGCAGTCCAGGCTGGGTGACAAAG----- 1147
Qy 1800 ATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1859
Db 1147 -----CAAAACACTGTCTCCAAAAAATTTAGGCTTGGCAAGGGCGCA 1189
Qy 1860 GTGGCTACGCGCTATAATCTCAGCAGCTCTGGGAGGCC-AAAGATGGAGGATTTGCTTGAAGC 1918
Db 1190 -CGGCTCAGCGCTGTGATCCCGAGCAGCTTTGGGAAGCCGAGCAGCAGATCAGTTGAGT 1248
Qy 1919 CAGGAGTTGGGACAGCCTGGGCAACATAGGGGATCCCAATCTCTACACACAAAAAAT 1978
Db 1249 CAGGAGTTGGGACAGCCTGGGCAACATAGGTGAACCCCTCTCTACTGAAAAATACAA- 1308
Qy 1979 TTTTAAATGAACAGGCAATTTGTCGATGCGCTATAGTCCAGCAGCAGCTCAAGAGGCACAG 2038
Db 1308 ----AAATAGCCGGTTGGTAGTGGTGGTGTGTAATCTTAGTACTTTGGGAGGCTGAG 1363
Qy 2039 GCGGAGGATCACTTGAGCCTGGGAGGTTGGTTGCGAGTGAGCTATGATTGTACCACTG 2098
Db 1364 GCAGGGGAATTCCTGAACCTGCGAGGCGGAGGCTGCGAGTGCAGCGGAGATTGCATCATG 1423
Qy 2099 CACTCCAGCCTGGGCAACAGAGAGACCTTGTCTCAAAAAATAACAAACTAAAAATTA 2158
Db 1424 CACTCTAGCCTGGCAACAGAGCTAGACTCCATCCCAAAAAAATAA----- 1472
Qy 2159 AAAGAAGACGAGAGATAGTGGGTGGTGGCTCACACCTGCAATCCAGCAGCTTTGGAA 2218
Db 1472 -----AAAGTAGCCGGGACGCTGGCTCAGGCTGTAAATCCAGCAGCTTTGGGA 1520
Qy 2219 GCGGAGGTGGGAGATCATCTGAGGCCAGGAGTTCAAGACCAGCCTGGCTTAACATGGT 2278
Db 1521 GCGGAGGCGGGCGGATCA--TGAGGCGAGGAGATCGAGACCATCTCTGGCTTAACACAGGTG 1578
Qy 2279 AATTCCTATCTTACCAAAAAATAAATAATAGCCAGGCGTGGTGGGACCTGTACT 2338
Db 1579 AATCCCTGTCTCTATAAATAACAAAAATTTAGCCCGGCGAGGCTGGGCGGCTGTAGT 1638
Qy 2339 GGGGAGGTGCCACCGAGCTACTGGGAGGCTGAGTCAGGAGAGATCGTTGAACCTGGGA 2398
Db 1639 -----CCAGCTACTCAGGAGAGTGAAGCAGGAGATGGCTGTAACCCGGG- 1695
Qy 2399 GCGGAGGTTGCGGTGAGTGTGGCAGCTGCACCTCCAGCCTGGGCAAGAGAGCGA 2458
Db 1685 GCGGAGCCTGCACTGAGCCGAGATCGGCCACTGCACCTCCAGCTTGGGTGACACCCGAGA 1744
Qy 2459 CTCTGTCTCCAAAAAAGAGAA 2481
Db 1745 CTCCGTCTCAAAAAAATAAATAA 1767


```

RESULT 5
PCT-US93-05857-6
; Sequence 6, Application PC/TUS9305857
; GENERAL INFORMATION:
; APPLICANT: Board of Regents
; APPLICANT: The University of Texas System
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DETECTING
; TITLE OF INVENTION: GENE REARRANGEMENTS AND TRANSLOCATIONS
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P. O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: USA
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/05857
; FILING DATE: 19930617
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/900,689
; FILING DATE: 17/06/92
; ATTORNEY/AGENT INFORMATION:
; NAME: Parker, David L.
; REGISTRATION NUMBER: 32,165
; REFERENCE/DOCKET NUMBER: ARCD:072/PAR
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512) 320-7200
; TELEFAX: (512) 474-7577
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8392 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
PCT-US93-05857-6

Query Match 6.6%; Score 324.6; DB 5; Length 8392;
Best Local Similarity 66.9%; Pred. No. 1.6e-50;
Matches 658; Conservative 0; Mismatches 229; Indels 96; Gaps 10;

QY 1520 GCCAGGCTGTGGCTCACGCTGTATCCAGCAGCTTTAAGAGCCGAGCGGATGGAT 1579
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 860 GCTGGCAGGCTGTGGCTCACGCTGTATCCAGCAGCTTTAAGAGCCGAGCGGAT 919
QY 1580 CAGTTGAGCCGAGGTTCAACACCCCTGAGCAACATGTTAAACCCCATCTCTACCA 1639
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 920 TGCTTGAGCCGAGGTTCAACACCCCTGAGCAACATGTTAAACCCCATCTCTATTT 979
QY 1640 AAAATACAAAAA-----TTAGCTGGCTGTGGCTGGCGCCCTGTA 1680
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 980 AAACAAAAAANAANAAGAAAGAAAGAGAGTTAGCAGGAGATGGTGCGATGTTA 1039
QY 1681 ATCCAGCTACTCAGGAGACTGAGCGAAGAGATCG-CTTGAACCTGGGAGCGAGAGTT 1739
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1040 GTCCAGGCTACTCAGGAGCTGAGATGAGAGGATGTTCTTGGAGCCAGGAATCAAGCT 1099
QY 1740 GCACTGAGCCGAGATTCGGCCACTGGAGCTCAGCCTCGGTGACAGAGAGCTGTCTAA 1799
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1100 GTAGTGAAGCTATGATTGACACTGCACTGAGTCCAGCTGGGTGACAAAG----- 1147
QY 1800 ATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1859
   || || || || || || || || || || || || || || || || || || || || ||

```

```

Db 1147 -----CAAAACACTGTCTCCAAAAAATTTAGGCTTGGCAAGCGCA 1189
QY 1860 GTGGCTCAGGCTATATCTCAGCAGCTCTGGAGGCC-AAGATGGAGGATGCTTTGAAGC 1918
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1190 -CGGCTCAGGCTGTGATCCAGCAGCTTTGGGAAGCGGAGCAGCAGATCACTTGAGCT 1248
QY 1919 CAGGAGTTTGGGACCAAGCTGGGCAACATAGGGGATCCCTCTCTACACACAAAAAAT 1978
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1249 CAGGAGTTTGGGACCAAGCTGGGCAACATAGGGGATCCCTCTCTCTCTCTCTCTCT 1308
QY 1979 TTTTATGAACCAAGCAGCTTTGGGATGGGCTATAGTCCAGCCACTCAAGAGGACAG 2038
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1308 ----AAATTAGCGGTTGGTAGTGGGTGCTGTATCTCTAGCTACTTGGGAGGCTGAG 1363
QY 2039 GCGGAGGATCACTTGGAGCTGGGAGTTGTGTGGTTCAGTATGATGTATGACACTG 2098
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1364 GCAGGGGAATTGCTCACTGCGAGGCGGAGCTGAGTCCAGTCCAGTCCAGTCCAGT 1423
QY 2099 CACTCAGGCTGGGCAACAGAGCAGACCTTGTCTCAAAAAAATAAACAATAAATAA 2158
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1424 CACTTAGCTGGCAACAGAGCTAGACTCCATCCAAAAAATAAATAAATAAATAA 1472
QY 2159 AAAAAGAGAGAGATAGTGGGTGGTGGCTCACACCTGCAATCCAGCAGCTTTGGAA 2218
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1472 -----AAAGTAGCGGCGAGCTGAGTCCAGTCCAGTCCAGTCCAGTCCAGT 1520
QY 2219 GCGGAGGTGGGAGATCACTGAGCCAGGAGTTTCAAGACCAAGCCTGGCTTAACATG 2278
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1521 GCGGAGGTGGGAGATCACTGAGCCAGGAGTTTCAAGACCAAGCCTGGCTTAACATG 1578
QY 2279 AATCTATCTTACCAAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 2338
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1579 AAACCTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1638
QY 2339 GGGGAGGTGCCACCAAGCAGTCTGAGGAGCTGAGTCCAGGAGTCCAGTCCAGTCC 2398
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1639 -----CCAGCTACTCAGGAGTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1685
QY 2399 GCGGAGGTGGGAGTGGGAGTGGGAGTGGGAGTGGGAGTGGGAGTGGGAGTGGGAG 2458
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1685 GCGGAGGTGGGAGTGGGAGTGGGAGTGGGAGTGGGAGTGGGAGTGGGAGTGGGAG 1744
QY 2459 CTCTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2481
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1745 CTCTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1767

```

```

RESULT 6
PCT-US95-07201-43/c
; Sequence 43, Application PC/TUS9507201
; GENERAL INFORMATION:
; APPLICANT: Chader, Gerald J.; Becerra, Sofia
; APPLICANT: Patricia; Schwartz, Joan P.;
; APPLICANT: Taniwaki, Takayuki
; TITLE OF INVENTION: PIGMENT EPITHELIUM
; TITLE OF INVENTION: DERIVED FACTOR: CHARACTERIZATION GENOMIC
; TITLE OF INVENTION: ORGANIZATION AND SEQUENCE OF THE PEDF GENE
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Morgan & Finnegan, L.L.P.
; STREET: 345 Park Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/07201
; FILING DATE: 06-JUN-1995

```

QY	1525	GCCTGGTGCCTCAGCCCTGTAAATCCCAGCACTTTTAAGAGGCCAAGGGCGATGGATCACTT	1584
Db	13774	GGCGAGTGGCTCAGCCTCAACCTATAATCCCAGCACTTTGGGAGGCTGAGGTGGGTGAATCACTT	13715
QY	1585	GAGCCGAGGAGTTCAACACACAGCCTGAGCACTGGTAAAAACC-CATCTCTACCAAAAA	1643
Db	13714	GAGTCAAGGAGTTGCGAGACCATCTGGGCACACCGTGAACCCGCTCTCTACTAAAAATA	13655
QY	1644	TACAAAAATTAGCTGGGCTTGGTGGCTGGCGCTGTAAATCCAGCTACTCAGGAGACTGA	1703
Db	13654	CAAAAAAATTAAACGGGCGATGGTGGTGCATGCCCTGTAAATCCAGCTACTCAGGAGGCTGA	13595
QY	1704	GGCAGAAGGATCGCTTGAACCTGGGAGGCGACAGGTTGCAGTGAGCCGAGATTGCGCCACT	1763
Db	13594	GGCAGAGATTCGCTTGCACCCAGAGGCGACAGGTTGCAGTGAGCTGAGATGATGCGCCACT	13535
QY	1764	GGACTCAGCGCTGGGTGACAGA--GAGCCTCTCTCTAAATTAATTAATTAATTTAA	1821
Db	13534	GTACTCAGCGCTGGGTGACAGAGTGAGACCCGTCTCAAAAAATAAATAAGTAATAATAA	13475
QY	1822	TTCAATTTTAAAAAGACGAAAGTGACGGCCAGGTGCAGTGGCTCAGCCCTAATATCTCA	1881
Db	13474	ATAAA--TATAAATATAAATTTGGAGCTAGGTGCAGTGGCTTATGCCCTATATCTCCA	13418
QY	1882	GCACCTCTGGGAGGCCAAGATGGA--GGATTGCTTGAAGCCAGGAGTTTGGGACCAGCCGTG	1940
Db	13417	GCACCTTTGGCAGGCCGAGGAGATGGAATCACTTGAGCTCAGGAGTTTCAGACCAACCTGA	13358
QY	1941	GCAACATAGGGGGATCCATCTCTACACAAAAAATTTTTTAAATGAACCAGGCACTGT	2000
Db	13357	GCAACATGGTGGGCCCCATCTCT--ACTAAAAATACCCAAAAAATTAGCTGGGCGTGGT	13300
QY	2001	GGCATGCCCTATAGTCCCAGCCACTCAAGAGGCACAGCGGGGAGGATCACTTGAGCGCTG	2060
Db	13299	GGACCTGCCCTGGTCCCAGCTATTTGGAGGCTGAGGCATGAGAATTTGGTTGAATCA	13240

Query Match	6.3%	Score 306.6;	DB 1;	Length 8174;
Best Local Similarity	74.7%	Pred. No. 3e-47;		
Matches 483;	Conservative	0;	Mismatches 144;	Indels 20;
Gaps 7;				

Qy	1519	GGCAGGCCTGGTGGCTACAGCCTGTAATCCACGACACTTTAAGAGGCCAAGCGCGATGGA	1578
Db	6955	GGCCAGCATGGTGCTCAGCGCGGTAATCCACGACACTGTGGAGGCCGAGGTGGCGAGA	6996
Qy	1579	TCACTTGAAGCCAGGAGTTCAACACGAGCCTGAGCAACATGGTAAACCCCATCTCTACC	1638
Db	6895	TCAC--GAGGTGAGGAGATCGAGAGCATCTCGTTAAACAGCATGAACCCCGCTCTACT	6938
Qy	1639	AAAAATAC--AAAAATTAGCTGGCGTTGGTGGCTGGCGCGCTGTAATCCAGACTACTACAG	1695
Db	6837	AAAAATACAAAAAAATTAAGCCGGCGGTGGTGGCGGGCGCGCTAGTATCCAGCTACTACAG	6778
Qy	1696	GAGACTGAGGCAGAGGATCGCTTGAACCTGGAGGCCAGAGGTTGCACTGAGCCGAGATT	1755

CITY: Atlanta
STATE: Georgia
COUNTRY: USA
ZIP: 30309-4530
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/273,411
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Pabst, Patrea L.
REGISTRATION NUMBER: 31,284
REFERENCE/DOCKET NUMBER: Wu106
TELECOMMUNICATION INFORMATION:
TELEPHONE: (404) 815-6508
TELEFAX: (404) 815-6555
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 3373 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
FEATURE:
NAME/KEY: misc.feature
LOCATION: 104..1201
OTHER INFORMATION: /note= "Nucleotides 104 thr
PUBLICATION INFORMATION:
AUTHORS: Larsen, et al.
JOURNAL: Proc. Nat'l Acad. Sci. USA
VOLUME: 87
PAGES: 6674-6678
DATE: 1990
RELEVANT RESIDUES IN SEQ ID NO: 2: FROM 1 TO 3
US-08-273-411-2

Query Match	6.3%;	Score	306.6;	DB 1;	Length	3373;		
Best Local Similarity	74.7%;	Pred.	No. 2.3e-47;					
Matches	483;	Conservative	0;	Mismatches	144;	Indels	20; Gaps	7;

Qy	1519	GGCCAGGCGCTGGTGCATCCGCCCTGTAACTCCACGCACTTTAAGAGGCCAACGGCGGATGGA	1578
Db	2373	GGCCAGGCGATGGTGCTTCACGCCGGTAACTCCACGCACTGTGGAGGCCGAGGTGGGCAGA	2314
Qy	1579	TCACTTGAGCCCGCAGAGTTCACACCACGCCCTGAGCAACATGGTAATAAACCCCATCTCTACC	1638
Db	2313	TCAC--GAGGTCAGGAGATCGAGACCATCTCTGTTACACGATGAACCCCCGCTCTCTACT	2256
Qy	1639	AAAAATAC--AAAATATAGCTGGGCTTGGTGGCTGGCGCCTGTAAATCCCAGCTACTCAG	1695
Db	2255	AAAAATACAAAAAAATTAGCGGCGCTGGTGGCGGGCGCTGTAGTCCCAGCTACTCAG	2196
Qy	1696	GAGACTGAGGAGAAAGATCGCTTGAACCTGGGAGGCGAGAGTTGCACTGAGCCGAGATT	1755
Db	2195	GAGGCTGAGGCAGA---ATGCGATGAACCCGGGAGGCAGAGCTTGCAGTAGCTGAGATC	2139
Qy	1756	GCGCACTGGACTCCAGCCTCGTGACAGAGCGCTGTCTCTAAATTAATTAATTA	1815
Db	2138	GCGCACTGCACCTGGCCTGGGCTAC--AGAGCAAGACTCCGCTCTCAATAAATAAATAA	2081
Qy	1816	ATTTAATTCAATTTTTAAAAAGACAAAAAGTCAGCGCCAGGTGCACTGCTCACGCCCTATA	1875
Db	2080	ATRAATAAATAATTAATAAA---AATATCCGGCTGGGCACAGTGGCTCATGCCAGTA	2025
Qy	1876	ATCTCAGCATCTGGGAGGCCAAGATGGAGATTCTTTGAAGCCAGGAGTTTGGGACCCAG	1935

Db	2024	ATCCAGCACATTTTGGGAGGCCAAGACACAGGTGGATGACA - AGGTCAGGAGATCGAGACCAT	1966
QY	1936	CCTGGCAACATAGGGGATCCCATCTTACACACAAAAAATTTTTTAATGAACACAGGC	1995
Db	1965	CCGGCCAAACATGGTGAACCCCTGCTCTAC - - - - - TAAAAATACAAAAATTAGCCNAGC	1911
QY	1996	ATTGTGSCATCGCGCTATAGTCCAGGCCATCAAGAGGCACAGCGGGAGGATCACTTGA	2055
Db	1910	ATAATGSCCTGTGCTGTAGTCCCATCTACTCAGGAGGCTGAGACAGGAGAATCGCTTGA	1851
QY	2056	GCCTGGGAGGTGTGGTTGCAGTGACGTATGATTGTACCACCTGCACCTCCAGCCTGGCAA	2115
Db	1850	ACACAGAGGCGGAAGTTGCAGTGGAGGCCAAGATCACGCCACTGCACCTCCAGCCTGGGCAA	1791
QY	2116	CAGACGAAGACCTTGTCTCAAAAATTAACAACACTAAATTTAAAAAA	2162
Db	1790	CAGAGCGAGACCCCTGTCTCCAAAAAATAAAAAAATAAAAAAATAAACCTAAAAAGA	1744

```

RESULT 10
US-08-525-058A-5/c
; Sequence 5, Application US/08525058A
; Patent No. 5770420
; GENERAL INFORMATION:
; APPLICANT: LOWE, JOHN B.
; TITLE OF INVENTION: METHODS AND PRODUCTS FOR THE SYNTHESIS
; OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,
; TITLE OF INVENTION: GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION
; OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCTURES
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLOH, SPIVAK, MCCLELLAND, MAIER & NEUSTADT, P.C.
; STREET: 1755 Jefferson Davis Highway, Fourth Floor
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/525,058A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Lavalleye, Jean-Paul M. P.
; REGISTRATION NUMBER: 31,451
; REFERENCE/DOCKET NUMBER: 2363-060-55
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)521-4500
; TELEFAX: (703)486-2347
; TELEX: 248855 OPAT UR
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8174 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ANTI-SENSE: NO
US-08-525-058A-5

```

	Query Match	6.3%	Score 306.6;	DB 3;	Length 8174;
	Best Local Similarity	74.7%;	Pred. No. 3e-47;		
	Matches 483;	Conservative	0;	Mismatches 144;	Indels 20; Gaps 7;
QY	1519	GGCCAGGCCCTGGTGGCTCAGCCCTGTAATCCACGACACTTTAAGAGGCCCAAGCCGATGGA	1578		
Db	6955	GGCCAGGCATGTTGGCTCAGCCGGTAAATCCAGCACTGTGGAGAGCCGAGTGGGCAGA	6896		

Sequence 3, Application PC/TUS9100899

GENERAL INFORMATION:

APPLICANT: Lowe, John B.

TITLE OF INVENTION: Method and Products For the Synthesis of

TITLE OF INVENTION: Oligosaccharide Structures on Glycoproteins, Glycolipids,

TITLE OF INVENTION: or as Free Molecules, and For the Isolation of Cloned

TITLE OF INVENTION: Genetic Sequences That Determine These Structur

NUMBER OF SEQUENCES: 16

CORRESPONDENCE ADDRESS:

ADDRESSEE: OBLON, SPIVAK, MCLELLAND, MAIER & NEUSTADT,

ADDRESSEE: P.C.

STREET: 1755 Jefferson Davis Highway, Suite 400

CITY: Arlington

STATE: Virginia

ZIP: 22202

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: PCT/US91/00899

FILING DATE: 19910214

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Lavalleye Ph.D., Jean-Paul

REGISTRATION NUMBER: 31,451

REFERENCE/DOCKET NUMBER: 2363-021-55 PCT

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703)521-5940

TELEFAX: (703)486-2347

TELEX: 248855 OPAT UR

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 8174 base pairs

TYPE: NUCLEIC ACID

STRANDEDNESS: unknown

TOPOLOGY: unknown

MOLECULE TYPE: cDNA

HYPOTHETICAL: NO

ORIGINAL SOURCE:

ORGANISM: Homo sapiens

TISSUE TYPE: Blood

FEATURE:

NAME/KEY: misc.feature

LOCATION: 4686..5780

OTHER INFORMATION: /label= mat_peptide

PCT-US91-00899-3

Query Match 6.3%; Score 306.6; DB 5; Length 8174;
Best Local Similarity 74.7%; Pred. No. 3e-47;
Matches 483; Conservative 0; Mismatches 144; Indels 20; Gaps 7;

QY 1519 GCCACAGCGCTGGTGCCTACCGCTGTATCCAGCAGCTTTAAGAGCCCAAGCGCGATGGA 1578

Db 6955 GCCACAGCGATGGTGGCTACCGCGGTGTATCCAGCAGCTGTGGAGCGCGAGTGGCGAGA 6896

QY 1579 TCACCTTGAGCCAGGAGTTCAACACAGCAGCTGAGCAACATGCTAAACCCCATCTCTACC 1638

Db 6895 TCAC--GAGTCAGGAGATCGAGACCATCTGTTGTAACACAGATGAACCCCGCTCTACT 6838

QY 1639 AAAAAATAC---AAAAATAGCTGGGCTGGTGGCTGGCGCTGTATCCAGCTACTCTCAG 1695

Db 6837 AAAAAATACAAAAAATAGCCGGCGGTGGTGGCGCGCTGTAGTCCAGCTACTCTAG 6778

QY 1696 GAGACTGAGGAGAGGATCGCTTGAACCTGGGAGCAGAGGTTGCAGTGAGCCGAGATT 1755

Db 6777 GAGGCTGAGGCAGA---ATGGCATGAACCCGGGAGCAGAGCTTGCAGTGAGCTGAGATC 6721

QY 1756 GGGCCACTGGACTCCAGCCTGGTGACAGAGAGCCTGTCTCTAAATTAATTAATTAATA 1815

Db 6720 GGGCCACTGCACTCTGGCTGGGCTAC--AGAGCAAGACTCCGCTCTCAATAAATAATAA 6663

QY 1816 ATTTAATCAATTTTAAAAAGACGAAGAAGTGCAGCCAGGTGCAGTGGCTCAGCGCTATA 1875
Db 6662 ATAAATAAATAATTAATAA---AAATATCCGGGCTGGGCACAGTGGCTCATGCCAGTA 6607
QY 1876 ATCTCAGCAGCTCTGGGAGCCCAAGATGGAGGATTCCTTGAAGCCAGGAGTTTGGGACCAG 1935
Db 6606 ATCCAGCAGCTCTGGGAGCCCAAGACAGTGGATGACA-AGTCAGGAGATCGAGACCAT 6548
QY 1936 CCTGGGCAACATAGGGGATCCCATCTCTACACACAAAAAATTTTTTAATGAACCCAGC 1995
Db 6547 CCGGCCCAACATGTGTGAACCCCTGTCTCTAC-----TAAAAATACAAAAAATTAGCCAGC 6493
QY 1996 ATTGTGGCATGGCTATATAGTCCAGCCACTCAAGAGCAGCAGCGGAGGATCACTTGA 2055
Db 6492 ATATGGCCCTGTGCTGTAGTCCCATCTACTCAGGAGGCTGAGACAGGAGATCGCTTGA 6433
QY 2056 GCCTGGGAGTTGTGGTTGCAGTGCAGTATGATTGTACACTGCAGCTCCAGCCTGGGCAA 2115
Db 6432 ACACAGGAGCGGAGTTGCAGTGCAGCAAGATCAGCCACTGCAGCTCCAGCCTGGGCAA 6373
QY 2116 CAGAGCAAGACCTGTCTCAAAAAATAAACAACAACTAAAAATTAATAA 2162
Db 6372 CAGAGCGAGCCCTGTCTCTCAAAAAAATAAACAACAACTAAAAAAGA 6326

RESULT 13

US-08-257-963B-9/c
Sequence 9, Application US/08257963B
Patent No. 5940686
GENERAL INFORMATION:
APPLICANT: Chader, Gerald J.; Becerra, S.
APPLICANT: Patricia; Schwartz, Joan P.;
APPLICANT: Taniwaki, Takayuki
TITLE OF INVENTION: PIGMENT EPITHELIUM
TITLE OF INVENTION: DERIVED FACTOR: CHARACTERIZATION OF ITS NOVEL
TITLE OF INVENTION: BIOLOGICAL ACTIVITY AND SEQUENCES ENCODING
TITLE OF INVENTION: AND EXPRESSING THE PROTEIN
NUMBER OF SEQUENCES: 42
CORRESPONDENCE ADDRESS:
ADDRESSEE: Morgan & Finnegan
STREET: 345 Park Avenue
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10154
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy Disk
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WORDPERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/257,963B
FILING DATE:
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/952,796
FILING DATE: 24-SEPT-1992
ATTORNEY/AGENT INFORMATION:
NAME: DOROTHY R. AUTH
REGISTRATION NUMBER: 36434
REFERENCE/DOCKET NUMBER: 20264126US1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 751-6849
TELEFAX: (212) 751-6849
INFORMATION FOR SEQ ID NO:
SEQUENCE CHARACTERISTICS:
LENGTH: 4421 Base Pairs
TYPE: Nucleic Acid
STRANDEDNESS: Double
TOPOLOGY: Unknown
MOLECULE TYPE: Genomic DNA
ORIGINAL SOURCE:

14
PCT-US95/07201-9/c
Sequence 9, Application PC/TUS9507201
GENERAL INFORMATION:
APPLICANT: Chader, Gerald J.; Becerra, Sofia
APPLICANT: Patricia; Schwartz, Joan P.;
APPLICANT: Taniwaki, Takayuki
TITLE OF INVENTION: PIGMENT EPITHELIUM
TITLE OF INVENTION: DERIVED FACTOR: CHARACTERIZATION GENOMIC
TITLE OF INVENTION: ORGANIZATION AND SEQUENCE OF THE PEDF GENE
NUMBER OF SEQUENCES: 43
CORRESPONDENCE ADDRESS:
ADDRESSEE: Morgan & Finnegan, L.L.P.
STREET: 345 Park Avenue
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10154
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy Disk
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WORDPERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/07201
FILING DATE: 06-JUN-1995
CLASSIFICATION:
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: 08/367,841
FILING DATE: 30-DEC-1994
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: 08/257,963
FILING DATE: 07-JUN-1994
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: 07/952,796
FILING DATE: 24-SEP-1992
ATTORNEY/AGENT INFORMATION:
NAME: DOROTHY R. AUTH
REGISTRATION NUMBER: 36434
REFERENCE/DOCKET NUMBER: 20264136PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 758-4800
TELEFAX: (212) 751-6849
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 4421 Base Pairs
TYPE: Nucleic Acid
STRANDEDNESS: Double
TOPOLOGY: Unknown
MOLECULE TYPE: Genomic DNA
ORIGINAL SOURCE:
ORGANISM: Human
FEATURE:
NAME/KEY: JT1
LOCATION:
IDENTIFICATION METHOD:
OTHER INFORMATION: 7.1 kb Bam HI
OTHER INFORMATION: fragment Derived from human placental

∞

```

RESULT 15
US-08-480-449-1/C
: Sequence 1, Application US/08480449
: Patent No. 5688927
: GENERAL INFORMATION:
: APPLICANT: Godiska, Ronald
: APPLICANT: Gray, Patrick W.
: TITLE OF INVENTION: MACROPHAGE DERIVED CHEMOKINE
: NUMBER OF SEQUENCES: 24
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
: STREET: 6300 Sears Tower, 233 South Wacker Drive
: CITY: Chicago
: STATE: Illinois
: COUNTRY: United States of America
: ZIP: 60606-6402
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/480,449
: FILING DATE:
: CLASSIFICATION: 530
: ATTORNEY/AGENT INFORMATION:
: NAME: Gass, David A.
: REGISTRATION NUMBER: 38,153
: REFERENCE/DOCKET NUMBER: 27866/32779
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 312/474-6300
: TELEFAX: 312/474-0448
: TELEX: 25-3856
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2923 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: CDNA
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 20..298
: US-08-480-449-1

Query Match 6.1%; Score 296.6; DB 2; Length 2923;
Best Local Similarity 71.3%; Pred. No. 1.4e-45;
Matches 463; Conservative 0; Mismatches 174; Indels 12; Gaps

QY 1507 AAGATAAGGAGGAGCCGCGCTGGTGGCTCAGCGCTGTAATCCCGAGCACCTTTAAGAGGCC 1566
Db 1831 AAAAAAGGGGAGGCTGGGTGTGATAGCTCAGCGCTGTAATCTCAGCACITTTGGGAGCAC 1772
QY 1567 AAGCGCGATGGATCACTTTAGCCCCAGGAGTTCAACACAGCGCTCAGCAACATGTTAAAC 1626
Db 1771 AAGGAAGGTGGATCACTTGAAGTTCAGGAGTTCAAGACGAGCTCGCCAAATATGTTGAAA - 1712
QY 1627 CCCATCTCTACCAAAAATAACAAAATTAGCTGGGCTTCGTCGGCTGGCGCTGTAATCCCA 1686
Db 1712 CCTGCTCTCTAAATAACAAAATTAGCCAGCGGTAGTGGCACACACCTGTGAATCCCA 1653
QY 1687 GCTACTCAGGAGCTGAGGCAGAGGATCGCTTGAACCTGGGAGGCAGAGTTCAGTGA 1746
Db 1652 GCTACTTTGGGAGCTGGGTGGGAGAACTACTTTGAACCCAGAGGTGAGGTTGTTAGTGA 1593

```

Qy	1507	ARGAT	AAGGAGAGCCAGCCCTGGTGGCTCAGCCTGTAACTCCAGCACATTTAAGAGCC	1566
Db	1831	AAAAA	GGGGGGCTGGGTGTATAGCTCAGCCTGTAATCTCAGCACTTTGGGAGCAC	1772
Qy	1567	AGCGCGATGGATCTTGACGCCAGGAGTTCAACACACAGCCTCAGCAACATGTTAAAC	1626	
Db	1771	ANGNAAGTGGATCACTTGAGTCAAGGTCAGGAGTTCAAGACAGCCTGGCCAAATATGTTGAAA	1712	
Qy	1627	CCCATCTCTACCAAAAATACAAAAATTAAGCTGGCTTCGTGGCTGGCGCTGTAAATCCCA	1686	
Db	1712	CTGTCTCTACTAATAAATAACAAAAATTAAGCAGCCGTAGTGGCACACACCTGTAAATCCCA	1653	
Qy	1687	GCTACTCAGGAGCTAGGCGAGAAGGATCGCTTGAACCTGGGAGGCAGAGGTTCCAGTGA	1746	
Db	1652	GCTACTTGGGAGCTGGGTGGGAGAACTCTTGAACCCAGAGGTCGAGGTTCTAGTGA	1593	


```

Qy 1747 GCCGAGATTGGCCACTGGACTCCAGCTCGGTGACAGAGAGCCCTGTCTCTAATAATTATT 1806
Db 1592 GCCGAGATCAGGCCACTGCACCTCCAGCTGGCGCAGACAGAGTGAGACCCCTGCCATAAAAAA 1533
Qy 1807 AATTAAATTAATTTAATTCAATTTTAAAAAGACGAAAGTGACGGCCAGGTGTCAGTGGCTC 1866
Db 1532 AAAAAAAGTTGGGGGGGAGTGGGGAGAGGGNAG----AGGCCAGGCACTGTGGCTC 1477
Qy 1867 ACGCCTATAATCTCAGCACTCTGGAGAGCCCAAGAT-GGAGGATTGCTTTGAAGCCAGGAGT 1925
Db 1476 ACGCCTGTAATCCCAAGCACTTTGGGAGSCTGAGGTGGCGGATTACCTGAGGACAGGAGT 1417
Qy 1926 TGGGACCAGCCTGGGCAACATAGGGGGATCCCATCTCTACACACAAAAAATTTTAA 1985
Db 1416 TCGAGACCAGCCTGGCCAAACATGTGTGAAGCCTCGTCTCTAC-----TAAAAATACAAAA 1362
Qy 1986 TGAACCGAGCATTTGGGCATGGCCTATAGTCCAGCCACTCAAGAGGCACAGCGGGAG 2045
Db 1361 TTAGTGGCGGTGTAGCAGGAACCTGTATCCAGCTACTCAGGAGCTGAGGCAAGAG 1302
Qy 2046 GATCACTTGAGCCTGGGAGGTTGTGGTTGCAGTGAGCTATGATTGTACCACTGCACCTCCA 2105
Db 1301 AATCACTTGAACCCAGGAGGAGGAGGCTTCAATAAGCCAGACCAACCATTTGCACCTCA 1242
Qy 2106 GCCTGGGCAAC-AGAGCAAGACCTTGTCTCAAAAAATAAACAACTAAAA 2153
Db 1241 GCGTGGGTGACAAGAGCGGAAACTCCATCTCAAAAAAATAAAAAA 1193

```

Search completed: October 31, 1999, 01:40:22
Job time: 2900 sec

This Page Blank (uspto)

Result No.	Score	Query		Length	DB	ID	Description
		Match	%				
1	1080	54.2	359	1	W69598		Mouse G-protein co
2	908.5	45.6	360	1	W69597		Mouse G-protein co
3	530	26.6	399	1	R68922		Murine C140 recept
4	530	26.6	395	1	W01954		Murine C140 recept
5	525	26.4	395	1	R66920		Murine C140 recept
6	522	26.2	395	1	W01952		Murine C140 recept
7	521.5	26.2	374	1	W51406		Human protease-act
8	508	25.5	398	1	R66921		Human C140 recepto
9	508	25.5	398	1	W01953		Human C140 recepto
10	507.5	25.5	369	1	W51405		Mouse protease-act
11	505.5	25.4	394	1	W51408		Human protease-act
12	494	24.8	397	1	R66923		Human C140 recepto
13	494	24.8	397	1	W01955		Human C140 recepto
14	473	23.8	425	1	R27240		Human thrombin rec
15	473	23.8	425	1	W51407		Human protease-act
16	472	23.7	425	1	R60698		Fragment of the hu
17	472	23.7	892	1	W16314		Human thrombin rec
18	378.5	19.0	322	1	R48715		G-protein coupled
19	378.5	19.0	322	1	W02687		G-protein coupled
20	374	18.8	328	1	R91225		Human placenta G-p
21	354.5	17.8	339	1	R53752		Seven transmembran
22	354.5	17.8	339	1	W07617		Human G-protein th
23	354.5	17.8	339	1	R48733		Human R12 seven tr
24	353.5	17.8	328	1	R91224		Mouse pancreas G-p
25	348	17.5	328	1	W09433		Human placenta pur
26	343	17.2	380	1	R53750		Seven transmembran
27	343	17.2	380	1	R48731		Human R20 seven tr
28	341.5	17.2	370	1	W62597		Human 7-transmembr
29	337.5	17.0	361	1	R54080		Epstein Barr virus
30	337.5	17.0	361	1	W53623		Epstein Barr virus
31	332.5	16.7	355	1	R28272		Sequence in a high
32	332.5	16.7	355	1	R80950		Recombinant high a
33	331.5	16.6	304	1	R48714		G-protein coupled
34	331.5	16.6	304	1	W02686		G-protein coupled
35	330.5	16.6	344	1	W04247		Human G-protein co
36	329.5	16.5	427	1	W61355		Human hypothalamic
37	325	16.3	365	1	W23606		Human p2y4 recepto
38	323.5	16.2	418	1	R39263		Human somatostatin
39	323.5	16.2	350	1	R68811		Interleukin-8 rece
40	323.5	16.2	1060	1	R70123		IL8-R type 1-GBP 1
41	323.5	16.2	350	1	R80756		Interleukin 8 rece
42	323.5	16.2	350	1	R80951		Recombinant high a
43	322.5	16.2	350	1	R27791		Interleukin-8 rece

Db 68 TGLKLTFLPVVYIIVFVIGLPSNGMALWIFLFRTRKKHPAVIYMANLADLLSVIWP 127
QY 131 PRIAYHLRGORWPEGAACRLATAALYGHMYGVSLLAAVSLDRYLALVHPL---RARAL 187
Db 128 LKISYHLHGNWYIGALCKVIGFYGNMYCSILFMTCLSVQRYWVWVNPMPGHPKRKA- 187
QY 188 RGRRLALGLCMAALMAALALPLTQRTFLRARSDRVLCHDALPLDAQASHWQPAFTC 247
Db 187 ---NIAVGSLAIWLLIFLVTIPLYVMKQTIYPALNITTCVDLPEEVLGDMFNFLS 243
QY 248 LALLGCFPL-LLAMLLCYGATLHTLAAG-----RRYGHALRLTAVVLASAVAFVPSN 300
Db 244 LA-IGVFLFPALLTASAYVLMIKTLRSSAMDEHSENKRQRAIRLIITVLMWYFICFRPSN 302
QY 301 LLLLLHYSDPSAWNLGAYVPSLALSTLNSCDVPFIYVYSAEFRDKVRAGLFQRS- 360
Db 303 LLLVHYFLIKTQROSHVYALYVALCULSTLNSCIDPFYVYFVSKDFRDHARNALLCRSV 362
QY 360 -PGDTVASKASAEGRSGMGTHSS 382
Db 363 RTVNRMQISLSSNKFESKSGSYSS 386

RESULT 6

W01952
ID W01952 standard; Protein: 395 AA.

AC W01952;

DT 01-APR-1997 (first entry)

DE Murine C140 receptor, including putative signal sequence.

KW C140 receptor; G-protein linked; coupled; seven pass; agonist;

KW antagonist; hypertension; hypotension; blood pressure.

OS Mus sp.

FH Key Location/Qualifiers

FT peptide 1..27

FT /note= "putative signal peptide, differs from

FT signal peptide encoded by a cDNA clone of

FT this receptor (see W01954), the signal

FT sequence given for the cDNA clone is

FT believed to be the correct sequence"

FT protein 28..395

FT /note= "mature protein"

FT modified_site 29

FT /note= "potential Asn-linked glycosylation site"

FT cleavage_site 34..35

FT /note= "putative protease receptor cleavage site"

FT region 78..100

FT /note= "transmembrane region I"

FT region 108..128

FT /note= "transmembrane region II"

FT region 148..169

FT /note= "transmembrane region III"

FT region 188..210

FT /note= "transmembrane region IV"

FT modified_site 220

FT /note= "potential Asn-linked glycosylation site"

FT region 244..264

FT /note= "transmembrane region V"

FT region 286..306

FT /note= "transmembrane region VI"

FT region 324..345

FT /note= "transmembrane region VII"

PN W09623225-A1.

PD 01-AUG-1996.

PF 25-JAN-1996; U01179.

PR 25-JAN-1995; US-390301.

PA (CORT-) COR THERAPEUTICS INC.

PI Scarborough RM, Sundell J;

DR WPI; 96-362813/36.

DR N-PSDB; T32036.

PT Vector for expression C140 cell surface receptor in host cell

PT useful to identify C140 agonist and antagonists, which are

PT antihypertensives and elevators of blood pressure, respectively

PS Example 1; Fig 1A-B; 60pp; English.

CC W01952 represents the murine C140 receptor (C140R), including a putative
CC signal peptide (see features table). DNA encoding C140R may be
CC engineered so as to allow the recombinant expression of C140R in a
CC suitable host cell, i.e. by removing the native expression-control
CC sequences and replacing them with control sequences operable in the host.
CC Such a recombinant receptor can be expressed on the surface of oocytes,
CC this provides a good assay system for identifying agonists/antagonists
CC of C140R. The C140 receptor is a G-protein linked receptor and a member
CC of the "seven-pass" transmembrane receptor superfamily (peptide chain
CC of the receptor passes through the cell membrane seven times, producing
CC seven transmembrane regions within the receptor molecule). The C140
CC receptor is involved in controlling blood pressure. C140 antagonists
CC (see W01942-W01951) are useful to inhibit signalling from this
CC receptor, resulting in an increase in blood pressure and are therefore
CC useful in pharmaceuticals for the treatment of hypotension (low blood
CC pressure). Conversely agonists (see W01914-W01941) of C140 are useful
CC in pharmaceuticals for the treatment of hypertension (high blood
CC pressure).
SQ Sequence 395 AA;

Query Match 26.2%; Score 522; DB 1; Length 395;

Best Local Similarity 34.1%; Pred. No. 5.4e-44;

Matches 131; Conservative 66; Mismatches 143; Indels 44; Gaps 11;

QY 13 GFSLSGGTQT-PSVYDESGSTGGDDSTPSILPAPRGYPGOVCANDSDTLELPDSSRALL 71

Db 33 GRSILGRLETPPI-----TGKG-----VPVEPGF-----SIDESASIL 67

QY 72 LGWVTRLPVLPALYGLVYVGLPANGALWV-LAQAPRLPSTMLNMLNATADLLALALP 130

Db 68 TGLKLTFLPVVYIIVFVIGLPSNGMALWIFLFRTRKKHPAVIYMANLADLLSVIWP 127

QY 131 PRIAYHLRGORWPEGAACRLATAALYGHMYGVSLLAAVSLDRYLALVHPL---RARAL 187

Db 128 LKISYHLHGNWYIGALCKVIGFYGNMYCSILFMTCLSVQRYWVWVNPMPGHPKRKA- 187

QY 188 RGRRLALGLCMAALMAALALPLTQRTFLRARSDRVLCHDALPLDAQASHWQPAFTC 247

Db 187 ---NIAVGSLAIWLLIFLVTIPLYVMKQTIYPALNITTCVDLPEEVLGDMFNFLS 243

QY 248 LALLGCFPL-LLAMLLCYGATLHTLAAG-----RRYGHALRLTAVVLASAVAFVPSN 300

Db 244 LA-IGVFLFPALLTASAYVLMIKTLRSSAMDEHSENKRQRAIRLIITVLMWYFICFRPSN 302

QY 301 LLLLLHYSDPSAWNLGAYVPSLALSTLNSCDVPFIYVYSAEFRDKVRAGLFQRS- 360

Db 303 LLLVHYFLIKTQROSHVYALYVALCULSTLNSCIDPFYVYFVSKDFRDHARNALLCRSV 362

QY 360 -PGDTVASKASAEGRSGMGTHSS 382

Db 363 RTVNRMQISLSSNKFESKSGSYSS 386

RESULT 7

W51406

ID W51406 standard; Protein: 374 AA.

AC W51406;

DT 12-OCT-1998 (first entry)

DE Human protease-activated receptor 3 (PAR3).

KW Protease-activated receptor 3; PAR3; thrombin receptor; mouse;

KW G-protein coupled receptor; agonist; antagonist; thrombosis;

KW atherosclerosis; restenosis; inflammation; blood coagulation;

KW blood clotting; heart attack; stroke; wound healing;

KW adult respiratory distress syndrome; glomerulosclerosis.

OS Homo sapiens.

FH Key Location/Qualifiers

FT modified_site 25..27

FT /note= "Asn is N-glycosylated"

FT cleavage_site 38..39

FT /note= "thrombin cleavage site"

FT peptide 48..51

FT /note= "hirudin-like sequence"

FT Modified_site 82..84 /note= "Asn is N-glycosylated"
FT Domain 95..117 /label= TM1
FT Domain 126..149 /note= "transmembrane domain 1"
FT Domain 168..191 /label= TM2
FT Domain 207..231 /note= "transmembrane domain 2"
FT Domain 261..286 /label= TM3
FT Domain 301..323 /note= "transmembrane domain 3"
FT Domain 331..333 /label= TM4
FT Domain 335..360 /note= "transmembrane domain 4"
FT Domain 361..386 /label= TM5
FT Domain 391..423 /note= "transmembrane domain 5"
FT Domain 431..433 /label= TM6
FT Domain 435..460 /note= "transmembrane domain 6"
FT Domain 461..486 /label= TM7
FT Domain 491..517 /note= "transmembrane domain 7"
FT WO9818456-A1.
PD 07-MAY-1998.
PF 29-OCT-1997; U19732.
PR 30-OCT-1996; US-742440.
PA (REGC) UNIV CALIFORNIA.
PI Connolly A, Coughlin SR, Ishihara H;
DR WPI; 98-271905/24.
DR N-PSDB; V07374.
PT DNA encoding protease-activated receptor 3 - for detection of
PT specific agonists and antagonists, potentially useful for treating
PT e.g. thrombosis, atherosclerosis, inflammation etc.
PS Claim 3; Page 41-42; 74pp; English.
CC This polypeptide comprises human protease-activated receptor 3
CC (PAR3), a cell surface protein which is specifically activated by
CC thrombin or a thrombin agonist, thereby activating signalling
CC events such as phosphoinositide hydrolysis, calcium ion efflux and
CC platelet aggregation. Its amino acid sequence was deduced from an
CC isolated cDNA clone (see V07372), and shows homology to human PAR1
CC and PAR2 (see W51407 and W51408). The mouse PAR3 amino acid
CC sequence (see W51405) is also provided. Also claimed are vectors,
CC host cells and an assay device. Host cells are used to screen
CC compounds for their ability to act as agonists or antagonists of
CC the effects of thrombin-PAR3 interaction. Agonists are used to
CC treat wounds, thrombosis, atherosclerosis, restenosis, inflammation
CC and other thrombin activated disorders. Antagonists (see W51415-21)
CC are used to control blood coagulation and thereby to treat heart
CC attack and stroke. They also mediate inflammatory and proliferative
CC responses to injury as occur in wound healing, atherosclerosis,
CC restenosis, pulmonary inflammation (ARDS) and glomerulosclerosis.
SQ Sequence 374 AA;

Query Match 26.2%; Score 521.5; DB 1; Length 374;
Best Local Similarity 33.8%; Pred. No. 5.7e-44;
Matches 120; Conservative 62; Mismatches 158; Indels 15; Gaps 5;

QY 3 GRLLWPLVLGFLSGGTQ-----TPSVYDESGSTGGDSTSTILPAPRGYGVQV----- 54
DB 10 GLLLLTFTFCOSGMENDTNLAKPTLPIKTRGAPPNSFEFP--FSALEGWTGATITVK 67
QY 54 --CANDSDT-LELPDSSRALLGHWVTRPLVPLVGLVGLPANGALWATQAPRLP 110
DB 68 IKCPESASHLVKNATWGLVTSLSKLIPALYLLVGVGVANATVLMWLFRTSIC 127
QY 111 STMLLMNLATADLLALPRLAYHLRGQWPFGEACRLATAALYGHMYSVLLAAV 170
DB 128 TTYFTNLATADFLFCVTLFKIAYHLNGNWNVGEVLCRAITTVIFGYGNMYSILLIACI 187

QY 171 SLDRIYALVHPLRARRLGRRLALGLCMAWLMARALPLTLQRTQFRLARSDRVLCMD 230
DB 188 SINRYLAIYHPTFYGRLPKHTYALVTCGLVWATVFLYMLPFFILKQEYLYVQPDITTC 247
QY 231 ALPLDAQASHWO-PAFTCLALLGCEPLPLAMLLCYGATLTLAASGRYRGHALRLTAVYL 289
DB 248 VHTNCSSSPFQLYYFISLAFFGLPFLVLIYCVAAIIRTNAYDHRWLVWYKASLLIL 307
QY 290 ASAVAFFVPSNLLHYSDPSPSAWNGLYGAYVPSLALSTLNSCVDPFIYYVVS 344
DB 308 VFTICFAPSNIILIIHHANYNNYNTDGLFYLYLALCLGSLNSCLDPFLYFLMS 362

RESULT 8
ID R66921 standard; Protein; 398 AA.
AC R66921;
DT 22-AUG-1995 (first entry)
DE Human C140 receptor.
KW G-protein-coupled receptor; G-protein; C140 receptor.
OS Homo sapiens.
FH Key Location/Qualifiers
FT protein 1..27
FT /label= signal peptide
FT modified_site 31
FT /label= Asn linked glycosylation site
FT modified_site 223
FT /label= Asn linked glycosylation site
FT cleavage_site 37..38
FT /label= protease receptor cleavage site
FT region 81..103
FT /label= transmembrane I
FT region 11..132
FT region 150..174
FT /label= transmembrane II
FT region 191..212
FT /label= transmembrane III
FT region 245..267
FT /label= transmembrane IV
FT region 289..309
FT /label= transmembrane V
FT region 327..348
FT /label= transmembrane VI
FT WO9503318-A.
PN 02-FEB-1995.
PD 26-JUL-1994; U08536.
PF 26-JUL-1993; US-097938.
PR (CORT-) COR THERAPEUTICS.
PA Scarborough RM, Sundelin J;
PI WPI; 95-075182/10.
DR N-PSDB; Q84558.
PT New DNA encoding recombinant C140 receptor - and novel agonists
PT and antagonists and specific antibodies with therapeutic and
PT diagnostic applications.
PT Disclosure; Fig 2; 57pp; English.
PS The availability of genomic DNA encoding the mouse protease
CC C140 receptor (see Q84557) permitted the retrieval of the corresp.
CC human gene. A human genomic library cloned in the vector EMBL3 was
CC screened using the entire coding region of the murine clone as a
CC probe. The recovered human gene including the DNA sequence and the
CC deduced AA sequence are shown in Q84558 & R66921. Subsequent
CC experiments indicated that the human C140 gene is located in the
CC same region of the long arm of chromosome number 5 (5q12-5q13),
CC as has been reported for the human thrombin receptor gene.
SQ Sequence 398 AA;

Query Match 25.5%; Score 508; DB 1; Length 398;
Best Local Similarity 36.5%; Pred. No. 1.4e-42;
Matches 110; Conservative 57; Mismatches 124; Indels 10; Gaps 5;

QY 67 SRALLGHWVTRPLVPLVGLVGLPANGALWV-LATQAPRLPSTMLMNLATADLL 125

```

Db 66 SASVLTGKLTTFVLPVYIVFVVGSLPSNGMALWFLFTKKKHPAVIYMANLADLLS 125
QY 126 ALALPPRIAYHLRGQWPFGAEACRLATAALYGHMYGVSLLLAANSLDRYLALVHPLRAR 185
Db 126 VWFPPLKIAIYHNGNWIYGEALCNVLIGFFYGNMYCSILFTCLSVQRYWVIVNPM-GH 184
QY 186 ALRGRLALGLCMAAWLMAALPLTLQRTQFRRLARSDRVLCCHDALPDAQASHWQPAF 245
Db 185 SRKKANIAIGISLAIWLLILLITPLVYVVKQTFIPALNITTCCHDVLPPQLLVGMFNYF 244
QY 246 TCLALLGCFL-PLLAMLLCYGATLHTLAASG-----RRYGHALRLTAVVLASAVAFVFP 298
Db 245 LSLA-IGVFLPAPLTASAYVIMRLRSSAMDENSEKKRKAIRKLIIVTVMYLICFTP 303
QY 299 SNLLLLHYSDSPSANGNLYGAYVPSLALSTLNSCDPFIYVYSAEFRDKVRAGLQF 358
Db 304 SNLLLVHYFLIKSQGSHVYALIVLCLSTLNSCIDPFYVYFVSHDFRDHAKNALLCR 363
QY 359 S 359
Db 364 S 364

```

```

RESULT 9
W01953
ID W01953 standard; Protein: 398 AA.
AC W01953;
DT 01-APR-1997 (first entry)
DE Human C140 receptor, with putative signal sequence.
KW C140 receptor; G-protein linked; coupled; seven pass; agonist;
OS antagonist; hypertension; hypotension; blood pressure.
FS Homo sapiens.
FH Key
FT peptide
FT 1. 27
FT /note= "putative signal peptide, differs from
FT signal peptide encoded by a cDNA clone of
FT this receptor (see W01953), the signal
FT sequence given for the cDNA clone is
FT believed to be the correct sequence"
FT 28. 398
FT /note= "mature protein"
FT protein
FT modified_site
FT /note= "potential Asn-linked glycosylation site"
FT cleavage_site
FT 37. 38
FT /note= "putative protease receptor cleavage site"
FT region
FT 81. 103
FT /note= "transmembrane region I"
FT region
FT 111. 132
FT /note= "transmembrane region II"
FT region
FT 151. 174
FT /note= "transmembrane region III"
FT region
FT 191. 212
FT /note= "transmembrane region IV"
FT modified_site
FT 223
FT /note= "potential Asn-linked glycosylation site"
FT region
FT 245. 267
FT /note= "transmembrane region V"
FT region
FT 289. 309
FT /note= "transmembrane region VI"
FT region
FT 327. 348
FT /note= "transmembrane region VII"
FT PN W09623225-A1.
PD 01-AUG-1996.
PF 25-JAN-1996; U01179.
PR 25-JAN-1995; US-390301.
PA (CORT-) COR THERAPEUTICS INC.
PI Scarborough RW, Sundelin J;
DR WPI; 96-362813/36.
DR N-PSDB; T32037.
DR Vector for expression C140 cell surface receptor in host cell -
PT useful to identify C140 agonist and antagonists, which are
PT antihypertensives and elevators of blood pressure, respectively

```

```

PS Example 2; Fig 2A-B; 60pp; English.
CC W01953 represents the human C140 receptor (C140R), including a putative
CC signal peptide (see features table). DNA encoding expression of C140R in a
CC engineered host cell, i.e. by removing the native expression-control
CC sequences and replacing them with control sequences operable in the host.
CC Such a recombinant receptor can be expressed on the surface of oocytes,
CC this provides a good assay system for identifying agonists/antagonists
CC of C140R. The C140 receptor is a G-protein linked receptor and a member
CC of the "seven-pass" transmembrane receptor superfamily (peptide chain
CC of the receptor passes through the cell membrane seven times, producing
CC seven transmembrane regions within the receptor molecule). The C140
CC receptor is involved in controlling blood pressure. C140 antagonists
CC (see W01942-W01951) are useful to inhibit signalling from this
CC receptor, resulting in an increase in blood pressure and are therefore
CC useful in pharmaceuticals for the treatment of hypotension (low blood
CC pressure). Conversely agonists (see W01914-W01941) of C140 are useful
CC in pharmaceuticals for the treatment of hypertension (high blood
CC pressure).
SQ Sequence 398 AA;

```

```

Query Match 25.5%; Score 508; DB 1; Length 398;
Best Local Similarity 36.5%; Pred. No. 1.4e-42;
Matches 110; Conservative 57; Mismatches 124; Indels 10; Gaps 5;
QY 67 SRALLGHWPTRLPALYGLVVLVPLANGALAV-LATQAPRLPSTMLMLNATADLLL 125
Db 66 SASVLTGKLTTFVLPVYIVFVVGSLPSNGMALWFLFTKKKHPAVIYMANLADLLS 125
QY 126 ALALPPRIAYHLRGQWPFGAEACRLATAALYGHMYGVSLLLAANSLDRYLALVHPLRAR 185
Db 126 VWFPPLKIAIYHNGNWIYGEALCNVLIGFFYGNMYCSILFTCLSVQRYWVIVNPM-GH 184
QY 186 ALRGRLALGLCMAAWLMAALPLTLQRTQFRRLARSDRVLCCHDALPDAQASHWQPAF 245
Db 185 SRKKANIAIGISLAIWLLILLITPLVYVVKQTFIPALNITTCCHDVLPPQLLVGMFNYF 244
QY 246 TCLALLGCFL-PLLAMLLCYGATLHTLAASG-----RRYGHALRLTAVVLASAVAFVFP 298
Db 245 LSLA-IGVFLPAPLTASAYVIMRLRSSAMDENSEKKRKAIRKLIIVTVMYLICFTP 303
QY 299 SNLLLLHYSDSPSANGNLYGAYVPSLALSTLNSCDPFIYVYSAEFRDKVRAGLQF 358
Db 304 SNLLLVHYFLIKSQGSHVYALIVLCLSTLNSCIDPFYVYFVSHDFRDHAKNALLCR 363
QY 359 S 359
Db 364 S 364

```

```

RESULT 10
W51405
ID W51405 standard; Protein: 369 AA.
AC W51405;
DT 12-OCT-1998 (first entry)
DE Mouse protease-activated receptor 3 (PAR3).
KW Protease-activated receptor 3; PAR3; thrombin receptor; mouse;
KW G-protein coupled receptor; agonist; antagonist; thrombosis;
KW atherosclerosis; restenosis; inflammation; blood coagulation;
KW blood clotting; heart attack; stroke; wound healing;
KW adult respiratory distress syndrome; glomerulosclerosis.
OS Mus sp.
PN W09818456-A1.
PD 07-MAY-1998.
PF 29-OCT-1997; U19732.
PR 30-OCT-1996; US-742440.
PA (REGC ) UNIV CALIFORNIA.
PI Connolly A, Coughlin SR, Ishihara H;
DR WPI; 98-271905/24.
DR N-PSDB; V07372.
DR DNA encoding protease-activated receptor 3 - for detection of
PT specific agonists and antagonists, potentially useful for treating

```


PT e.g. thrombosis, atherosclerosis, inflammation etc.
PS Claim 3; Page 39-40; 74pp; English.
CC This polypeptide comprises mouse protease-activated receptor 3
CC (PAR3), a cell surface protein which is specifically activated by
CC thrombin or a thrombin agonist, thereby activating signalling
CC events such as phosphoinositide hydrolysis, calcium ion efflux and
CC platelet aggregation. Its amino acid sequence was deduced from an
CC isolated cDNA clone (see V07372). The human PAR3 amino acid
CC sequence (see W51406) is also provided. Also claimed are vectors,
CC host cells and an assay device. Host cells are used to screen
CC compounds for their ability to act as agonists or antagonists of
CC the effects of thrombin-PAR3 interaction. Agonists are used to
CC treat wounds, thrombosis, atherosclerosis, restenosis, inflammation
CC and other thrombin activated disorders. Antagonists (see W51415-21)
CC are used to control blood coagulation and thereby to treat heart
CC attack and stroke. They also mediate inflammatory and proliferative
CC responses to injury as occur in wound healing, atherosclerosis,
CC restenosis, pulmonary inflammation (ARDS) and glomerulosclerosis.
SQ Sequence 369 AA;

Query Match 25.5%; Score 507.5; DB 1; Length 369;
Best Local Similarity 31.4%; Pred. No. 1.4e-42;
Matches 111; Conservative 76; Mismatches 151; Indels 15; Gaps 6;
QY 5 LLLNPLVL---GFLSGGTQTPSVYDESGTGGDDSTPSILPAP--RGYPGV----- 54
DB 11 LFLFVTVQCGINVSNSAKPTIK--SFNGQPPTFEFFLSDIEGWTGATTIKAE 68
QY 54 CANDS-DTLELPDSSRALLGQWTRVLPALYGLVLYVGLPANGALWLAQAPRLPST 112
DB 69 CPESISTLHVNNATIGYLRSSLTQVIPAIIYLLFVGVPSNVITLWKLRTKSLV 128
QY 113 MLNMLNATADLLALALPPIAYHLRGQWPFGEAARLATAALYGHMYSVLLAAVSL 172
DB 129 IFHTNLATADLFCVTLPEKIAHNGNWNVEGEVCMRTTTFVFGNYCAILLTCGI 188
QY 173 DRYALVHPLPARALGRRLALGCMALWMAALAPLTQLQRTFLARSDRVLCHDAL 232
DB 189 NRYLATAHPFTYQKLPRKFSFLSGIWMVFLYMLPFVILKQEHVHSEITTCDDV 248
QY 233 -PLDAQASHMOPACTLALLCFLPLAMLCYCATLHTLAASGRYRGHARLTAVVLAS 291
DB 249 DACSPSSFRYYVSLAFFGLFPFVFIICYTTLHLKSKDRWLGKAVILLIYI 308
QY 292 AVAFVPSNLLLHYDPSANGNLYGAYVPSLALSTLNSCVDPFIYYVVS 344
DB 309 FTICFAPNTIILVIHANYHHNTDSLYFMYLIALCLGSLNSCLDPLFLYFVMS 361

RESULT 11
W51408
ID W51408 standard; Protein; 394 AA.
AC W51408;
DT 12-OCT-1998 (first entry)
DE Human protease-activated receptor 2 (PAR2).
KW Protease-activated receptor 2; PAR2; PAR3; thrombin receptor;
KW human.
OS Homo sapiens.
FH Key Location/Qualifiers
FT Cleavage_site 36..37 /note= "thrombin cleavage site"
FT
PN W09818456-A1.
PD 07-MAY-1998.
PF 29-OCT-1997; U19732.
PR 30-OCT-1996; US-742440.
PA (REGC) UNIV CALIFORNIA.
PI Connolly A, Coughlin SR, Ishihara H;
DR WPI; 98-271905/24.
PT DNA encoding protease-activated receptor 3 - for detection of
PT specific agonists and antagonists, potentially useful for treating
PT e.g. thrombosis, atherosclerosis, inflammation etc.
PS Example 1; Page 43-44; 74pp; English.

CC This polypeptide comprises human protease-activated receptor 2
CC (PAR2). The physiological activator of PAR2 remains unknown;
CC it is not activated by thrombin. The invention relates to novel
CC mouse and human PAR3 (see W51405-06) that show homology to PAR2 and
CC which are specific receptors for thrombin. They can be used to
CC screen for specific agonists and antagonists of thrombin useful
CC e.g. for treating atherosclerosis, thrombosis and inflammation.
SQ Sequence 394 AA;

Query Match 25.4%; Score 505.5; DB 1; Length 394;
Best Local Similarity 34.1%; Pred. No. 2.4e-42;
Matches 122; Conservative 64; Mismatches 151; Indels 21; Gaps 9;
QY 10 LVLGFSLSGGTQTPSVYDESGTGGDDSTPSILPAPRGYPGVQVCANDSTLELPDSSRA 69
DB 16 LAASLSCSTQGTNRSSKRLGKVDGTSHV--TGKGVTVET-----VFSVDEFSAS 67
QY 70 LLLGWPTRLVPALYGLVLYVGLPANGALWV-LAQAPRLPSTMLNMLNATADLLALA 128
DB 68 VLTGKLTVPFPIYITIVFVGLPSNGMALWVLFRTKKKHPAVIYMANLALDLSVIV 127
QY 129 LPPIAYHLRGQWPFGEAARLATAALYGHMYSVLLAAVSLDRYLALVHPLRARALR 188
DB 128 FPLKIAHYHNGNWIYGEALCNVLIGFFYGNMYCSILFMTCLSVQRVWVIVNPM-GHSRK 186
QY 189 GRRALGLCMAWMAALAPLTQLQRTFLARSDRVLCHDALPLDAQASHMOPACTL 248
DB 187 KANIAIGISLAIWLLILLVTPLYVVKQTFIPALNTTCHDVLV--EQLLVGDP-FLSL 243
QY 249 ALLGCFL-PLLAMLLCGATLHTLAASG-----RRYGHALRLTAVVLASAVAFVPSNL 301
DB 244 A-IGVFPFAFLASAYVLMIRLSAMDENSEKKRRAIKLIVTVLAWLYLCFTFPSNL 302
QY 302 LLLHYSDPSANGNLYGAYVPSLALSTLNSCVDPFIYYVSAEFKDKVRAGLFQRS 359
DB 303 LLVVHYFLIKSQGSHYALYVALCLSTLNSCIDPFVYFVSHDFDHAKNALLCRS 360

RESULT 12
R66923
ID R66923 standard; Protein; 397 AA.
AC R66923;
DT 22-AUG-1995 (first entry)
DE Human C140 receptor encoded by cDNA.
KW G-protein-coupled receptor; G-protein; C140 receptor.
OS Homo sapiens.
PN W09503318-A.
PD 02-FEB-1995.
PF 26-JUL-1994; U08536.
PR 28-JUL-1993; US-097938.
PA (CORT-) COR THERAPEUTICS.
PI Scarborough RM, Sundelin J;
DR WPI; 95-075182/10.
DR N-PSDB; Q84560.
PT New DNA encoding recombinant C140 receptor - and novel agonists
PT and antagonists and specific antibodies with therapeutic and
PT diagnostic applications.
PS Example; Fig 11; 57pp; English.
CC A human intestinal tumour cDNA library was subjected to PCR using
CC primers designed from the genomic clone (see Q84558) and the
CC amplified fragment was cloned in pSG5 and sequenced. There are
CC four AA differences between the cDNA encoded sequence and that
CC encoded by the genomic DNA. The genomic DNA sequence and deduced
CC AA sequence are given in Q84560 & R66923.
SQ Sequence 397 AA;

Query Match 24.8%; Score 494; DB 1; Length 397;
Best Local Similarity 32.7%; Pred. No. 3.4e-41;
Matches 117; Conservative 64; Mismatches 159; Indels 18; Gaps 7;
QY 10 LVLGFSLSGGTQTPSVYDESGTGGDDSTPSILPAPRGYPGVQVCANDSTLELPDSSRA 69

```
Db 16 LAASLSCSGTIOGTRSSKGRSLIGKVDGTSVH--TGKGVTVET-----VFSVDEFSAS 67
QY 70 LLIGWTPTRLVPAALYGLVVLVGLPANGALWV-LATQAPRLPSTMLMNLATADLLALA 128
Db 68 VLAKGTTTFLPITVYIVFVAVGLPSNGMALWFLFTKKKHPAVIYMANLADLLSVIW 127
QY 129 LPPIAYHLRGQRPGEAAACRLATAALYGHMYGVSLLAAVSLDRYLAALVHPLRARALR 188
Db 128 FPKIAYHGHNNWIYGEALCNVLIGFFYRNMYCSILFTCLSVQRYWVIVNPM-GHSRK 186
QY 189 GRLALGLCMAAWLMAALALPLTLQRTFLARSDRVLCHDALPDAQASHWQPAFTCL 248
Db 187 KANIAIGISLAIWLLTLLVTPLYVVKQTIFIPALNITTCDDVLPQOLLVGMDFNLSL 246
QY 249 ALLGCFL-PLLAMLLCYGATLHTLAAG-----RRYGHALRLTAVVLASAVAFVPSNL 301
Db 247 A-IGVELFPAFLTASAYVLMIRMLRSSAMDENSEKRRKRAIKLIVTVLGMYLICFTPSNL 305
QY 302 LLLHYSDPSAWGNLYGAYVPSLALSTLNSCVDPFIYVYVSAEFRDKVRAGLFORS 359
Db 306 LLVWHFLIKSQGSHVYALYVALCLSTLNSCIDPFYIFVSHDFRDHAKNALLCRS 363

RESULT 13
W01955
ID W01955 standard; Protein; 397 AA.
AC W01955;
DT 02-APR-1997 (first entry)
DE Human C140 receptor.
KW C140 receptor; G-protein linked; coupled; seven pass; agonist;
KW antagonist; hypertension; hypotension; blood pressure.
OS Homo sapiens.
FH Key Location/Qualifiers
FT Peptide
FT 1..27 /note= "the signal peptide differs from that
FT encoded by a genomic DNA sequence for
FT this receptor (see W01953), the signal
FT sequence given here is believed to be
FT the correct sequence"
FT protein
FT 28..397
FT /note= "mature protein"
PN W09623225-Al.
PD 01-AUG-1996.
PF 25-JAN-1996; U01179.
PR 25-JAN-1995; US-390301.
PA (CORT-) COR THERAPEUTICS INC.
PI Scarborough RM, Sundellin J.
DR WPI: 96-362813/36.
DR N-PSDB; T32039.
PT Vector for expression C140 cell surface receptor in host cell -
PT useful to identify C140 agonist and antagonists, which are
PT antihypertensives and elevators of blood pressure, respectively
PS Example 5; Fig 11A-B; 60pp; English.
CC W01953 represents the human C140 receptor (C140R). DNA encoding C140R
CC may be engineered so as to allow the recombinant expression of C140R in
CC a suitable host cell, i.e. by removing the native expression-control
CC sequences and replacing them with control sequences operable in the host.
CC Such a recombinant receptor can be expressed on the surface of oocytes,
CC this provides a good assay system for identifying agonists/antagonists
CC of C140R. The C140 receptor is a G-protein linked receptor and a member
CC of the "seven-pass" transmembrane receptor superfamily (peptide chain
CC of the receptor passes through the cell membrane seven times, producing
CC seven transmembrane regions within the receptor molecule). The C140
CC receptor is involved in controlling blood pressure. C140 antagonists
CC (see W01942-W01951) are useful to inhibit signalling from this
CC receptor, resulting in an increase in blood pressure and are therefore
CC useful in pharmaceuticals for the treatment of hypotension (low blood
CC pressure). Conversely agonists (see W01914-W01941) of C140 are useful
CC in pharmaceuticals for the treatment of hypertension (high blood
CC pressure). 397 AA;
SQ Sequence
```

```
Query Match 24.8%; Score 494; DB 1: Length 397;
Best Local Similarity 32.7%; Pred. No. 3.4e-41;
Matches 117; Conservative 64; Mismatches 159; Indels 18; Gaps 7;
QY 10 LVLGFLSGGTQTPSVYDESGTGGDDSTPSILPAPRGYPQOVCAANDSTLELPDSSRA 69
Db 16 LAASLSCSGTIOGTRSSKGRSLIGKVDGTSVH--TGKGVTVET-----VFSVDEFSAS 67
QY 70 LLIGWTPTRLVPAALYGLVVLVGLPANGALWV-LATQAPRLPSTMLMNLATADLLALA 128
Db 68 VLAKGTTTFLPITVYIVFVAVGLPSNGMALWFLFTKKKHPAVIYMANLADLLSVIW 127
QY 129 LPPIAYHLRGQRPGEAAACRLATAALYGHMYGVSLLAAVSLDRYLAALVHPLRARALR 188
Db 128 FPKIAYHGHNNWIYGEALCNVLIGFFYRNMYCSILFTCLSVQRYWVIVNPM-GHSRK 186
QY 189 GRLALGLCMAAWLMAALALPLTLQRTFLARSDRVLCHDALPDAQASHWQPAFTCL 248
Db 187 KANIAIGISLAIWLLTLLVTPLYVVKQTIFIPALNITTCDDVLPQOLLVGMDFNLSL 246
QY 249 ALLGCFL-PLLAMLLCYGATLHTLAAG-----RRYGHALRLTAVVLASAVAFVPSNL 301
Db 247 A-IGVELFPAFLTASAYVLMIRMLRSSAMDENSEKRRKRAIKLIVTVLGMYLICFTPSNL 305
QY 302 LLLHYSDPSAWGNLYGAYVPSLALSTLNSCVDPFIYVYVSAEFRDKVRAGLFORS 359
Db 306 LLVWHFLIKSQGSHVYALYVALCLSTLNSCIDPFYIFVSHDFRDHAKNALLCRS 363

RESULT 14
R27240
ID R27240 standard; Protein; 425 AA.
AC R27240;
DT 25-MAY-1994 (second entry)
DT 15-FEB-1993 (first entry)
DE Human thrombin receptor (TR).
KW Diagnosis; cardiovascular disease; wound healing; restenosis;
KW thrombosis; unstable angina treatment; myocardial infarction;
KW thrombotic; thromboembolytic stroke.
OS Homo sapiens.
FH Key Location/Qualifiers
FT Peptide
FT 1..26 /label= sig_peptide
FT 28..425 /label= mat_protein
FT 28..41 /note= "activation peptide RPESKATNATLDPR"
FT cleavage_site 41..42
FT /note= "thrombin-catalysed cleavage-site"
FT disulfide_bond 175..254
FT /note= "analogous to rhodopsin and beta-2
FT adrenergic receptor"
FT domain 100..129
FT /note= "transmembrane domain I"
FT domain 137..160
FT /note= "transmembrane domain II"
FT domain 177..197
FT /note= "transmembrane domain III"
FT domain 219..239
FT /note= "transmembrane domain IV"
FT domain 266..294
FT /note= "transmembrane domain V"
FT domain 311..335
FT /note= "transmembrane domain VI"
FT domain 348..367
FT /note= "transmembrane domain VII"
FT region 34..36
FT /note= "consensus N-linked glycosylation site"
FT region 62..64
FT /note= "consensus N-linked glycosylation site"
FT region 75..77
FT /note= "consensus N-linked glycosylation site"
```

```

region      250..252
FT          /note= "consensus N-linked glycosylation site"
FT          259..261
region      /note= "consensus N-linked glycosylation site"

W09214750-A.
PN
PD 03-SEP-1992.
PF 19-FEB-1992: U01312.
PR 19-FEB-1991: US-637769.
PR (07-NOV-1991: US-789184.
PA (CORT-) COR THERAPEUTICS INC.
PA (REGC ) UNIV CALIFORNIA.
PI Coughlin SR, Scarborough RM;
PI WPI: 92-316119/38.
DR N-PSDB; Q28568.
PT DNA encoding cell surface receptor for thrombin - useful for
PT determining thrombin in diagnosing e.g. cardiovascular diseases,
PT also to treat wound healing, restenosis etc.
PS Disclosure; Fig 1: 81pp; English.
CC The TR DNA sequence can be used in the prepn. of diagnostics to
CC determine thrombin levels in samples, and screening tools for
CC candidate substances which affect thrombin activity in vivo.
CC Thrombosis may be diagnosed in a mammal by measuring the presence,
CC absence or amt. of the cleaved activation peptide of the TR.
SQ Sequence 425 AA;

Query Match      23.8%; Score 473; DB 1; Length 425;
Best Local Similarity 32.1%; Pred. No. 4.6e-39;
Matches 133; Conservative 76; Mismatches 169; Indels 36; Gaps

QY 4 RLLMLPLVLGFSLSG---GTQTPSVYDESGTGSGDDSTPSILPAP-----RG 48
   ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db 5 RLLL--VAACFSLCGPLLSARTARRPESKATNATLDRSFLLRNPNDKYEFWEDEKN 52
   ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
QY 49 YPG-----QVCANDSTL--ELP-----DSSRALLGWPTRIPLVPAlyGLVLVWGLPANG 96
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 63 ESGLTEYRLVSIKSSPLQKLPAFISDASGYLTSSWL--TLFVPSVYGVFVWSLPLNI 121
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
QY 97 LAIWLV-LATQAPRLPTMLLNATLADLLALAPRIAYHLRGORWPTGEAACRLATAA 155
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 122 MAIVFVFLKMKVKRPVVYMLHATADVLVSVLPKISYIFSGSDWQFSGELCRFVTAA 181
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
QY 156 LYGHMYGVSLLLAASLDRLYLALVHLPRALARALRGRLALGLCMAAWLMAAALPILQOR 215
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 182 FYCNMYASILLMTVISIDRELAVVYPMQSLSWRTLGRASFTCLAIWALAIGVWPLVKE 241
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
QY 216 QTFRLARSDVLCHDALPLDAQASHQWPAFTCLALGCGFLPLMLLCLYCATILHTLAAG 275
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 242 QTQVPGNLNTTCHDVLNFTLEGYIAYYFSAFSAVFFVPLIISTVCVYSIRCLSSA 301
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
QY 276 ----RRYGHALRLTAVVLASAVAFFVPSNLLLLHSDPS-PSAWGNLYGAVVPSLALST 330
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 302 VANESKSRALFLSAAVFCIFICFTGNTVLLIAHVSFLSHSTTEAYFAVILCVYSS 361
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
QY 331 LNSCVDPFIYIYVSAERDKVRAGLQFQSPGDTVASKASAEGRSGMGTHSLL 384
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 362 ISSCIDPLIYVYASRCQRYVGIILCKSSDPSSYNSQGLMASKMDYCSNL 415
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :

```

This Page Blank (uspto)

Result No.	Query				DB	ID	Description
	Score	Match	Length				
1	1991	100.0	385	4	076067	O76067 homo sapien	
2	1466	73.6	396	11	088634	O88634 mus musculus	
3	357	17.9	328	13	093361	O93361 meleagris g	
4	353	17.4	374	13	057466	O57466 meleagris g	
5	345.5	17.4	370	4	015132	O15132 homo sapien	
6	344	17.3	380	6	097666	O97666 macaca mulla	
7	340.5	17.1	345	11	070129	O70129 cavia porce	
8	336	16.9	371	11	092280	O92280 mus musculus	
9	332	16.7	361	11	035611	O35611 rattus norv	
10	321.5	16.1	371	11	088854	O88854 mus musculus	
11	320	16.1	387	4	043603	O43603 homo sapien	
12	317.5	15.9	370	11	088626	O88626 rattus norv	
13	314	15.8	342	13	093239	O93239 cyprinus ca	
14	313	15.7	368	4	060755	O60755 homo sapien	
15	309.5	15.5	337	4	075819	O75819 homo sapien	
16	309.5	15.5	370	11	054914	O54914 rattus norv	
17	308.5	15.5	353	13	079960	P79960 xenopus lae	
18	306.5	15.4	370	11	088853	O88853 mus musculus	
19	304	15.3	371	4	043190	O43190 homo sapien	
20	300.5	15.1	367	11	088410	O88410 mus musculus	
21	300	15.1	347	11	088535	O88535 mus musculus	
22	298.5	15.0	355	6	097664	O97664 macaca mulla	
23	298.5	15.0	383	12	089609	Q89609 equine herp	
24	298	15.0	352	6	018770	O18770 pan troglod	
25	297.5	14.9	356	6	097571	O97571 canis famli	
26	297.5	14.9	475	11	088680	O88680 cavia porce	
27	297	14.9	352	6	077776	O77776 cercocobus	
28	297	14.9	352	6	097975	O97975 macaca arct	
29	294.5	14.8	358	11	092213	Q92213 cavia porce	

	Query Match	100.0%;	Score 1991;	DB 4;	Length 385;	
	Best Local Similarity	100.0%;	Pred. No. 1.8e-142;			
	Matches 385;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;	
QY	1	MNGRLLWPVLVGLFSLGGGTQTSPSYDSEGSTGGDDSTPSILPAPRGYPGVCA N SDDT	60			
Db	1	MNGRLLWPVLVGLFSLGGGTQTSPSYDSEGSTGGDDSTPSILPAPRGYPGVCA N SDDT	60			
QY	61	LELPDSSRALLLGWPTRLVPALYGLVVVGLPANGIALWVLATQAPRLPSTMLLN L AT	120			
Db	61	LELPDSSRALLLGWPTRLVPALYGLVVVGLPANGIALWVLATQAPRLPSTMLLN L AT	120			

```
QY 121 ADLLALALPPRIAYHLRGWPFGEAACLATAAALYGHMYGVLAAVSLDRYLALVH 180
|||
DB 121 ADLLALALPPRIAYHLRGWPFGEAACLATAAALYGHMYGVLAAVSLDRYLALVH 180
QY 181 PLRALARGRRLALGLCMAAWLMAALALPLTLQRTFLRSLDRVLCVLDALPLDQAASH 240
|||
DB 181 PLRALARGRRLALGLCMAAWLMAALALPLTLQRTFLRSLDRVLCVLDALPLDQAASH 240
QY 241 WPAFTCLALLCGFLPLALLMLCYGATLTLAASGRYGHALRLTAVLASAVAFVPSN 300
|||
DB 241 WPAFTCLALLCGFLPLALLMLCYGATLTLAASGRYGHALRLTAVLASAVAFVPSN 300
QY 301 LLLLHYSDPSAMGNLYGAYVPSLALSTLNSCVDPFYIYVSAEFRKVRAGLFQSP 360
|||
DB 301 LLLLHYSDPSAMGNLYGAYVPSLALSTLNSCVDPFYIYVSAEFRKVRAGLFQSP 360
QY 361 GDTVASKASAEGRSGMGTHSSLLQ 385
|||
DB 361 GDTVASKASAEGRSGMGTHSSLLQ 385

RESULT 2
O88634 PRELIMINARY; PRT; 396 AA.
AC O88634;
DT 01-NOV-1998 (Tremblrel. 08, Created)
DT 01-NOV-1998 (Tremblrel. 08, Last sequence update)
DT 01-MAY-1999 (Tremblrel. 10, Last annotation update)
DE PROTEASE-ACTIVATED RECEPTOR 4.
GN PAR4.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RA KAHN M.L., HAMMES S.R., BOTKA C., COUGHLIN S.R.;
RT "Gene and locus structure and chromosomal localization of the
RT protease-activated receptor gene family.";
RL J. Biol. Chem. 0:0-0(1998).
RN [2]
RP SEQUENCE FROM N.A.
RA KAHN M.L., ZHENG Y.W., HUANG W., BIGORNIA V., ZENG D., MOFF S.,
RA FARESE R.V., TAM C., COUGHLIN S.R.;
RT "A dual thrombin receptor system for platelet activation.";
RL Nature 0:0-0(1998).
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
DR EMBL; AF080215; AAC28861.1; -.
DR PFAM; PF00001; 7tm1; 1.
KW PROSITE; PS00237; G-PROTEIN RECEPTOR; 1.
KW PROSITE; G-protein coupled receptor; Transmembrane; Glycoprotein.
SQ SEQUENCE 396 AA; 42703 MW; 23C7AF73 CRC32;
```

Query Match 73.6%; Score 1466; DB 11; Length 396;
Best Local Similarity 74.5%; Pred. No. 5.3e-103;
Matches 292; Conservative 31; Mismatches 55; Indels 14; Gaps 4;

```
QY 6 LLWPLVGLSLS--GGTQTPSYVDESGTGGDDST--PSI-----LPAPRGYPQ 52
|||
DB 5 LLYPLVLGLSLSLAEGIQTPSYVDESGTGGDDST--PSI-----LPAPRGYPQ 54
|||
QY 53 VCANDSDTLELPDSSRALLGWPTRLVPLVGLVPLANGALVLAQAPRLPST 112
|||
DB 65 FCANDSDTLELPASSALLGWVSTRLVPLVGLVPLANGALVLAQAPRLPST 124
|||
QY 113 MLMLNLTADLLALPLPRIAYHLRGWPFGEAACLATAAALYGHMYGVLAAVSL 172
|||
DB 125 ILLTNLAVADSLALVPPPLAYHLRGWPFGEAACLATAAALYGHMYGVLAAVSL 184
|||
QY 173 DRYLALVHLPLRALGRRLALGLCMAAWLMAALALPLTLQRTFLRSLDRVLCVLDAL 232
|||
```

```
DB 185 DRYLALVHLPLRALGRRLALGLCMAAWLMAALALPLTLQRTFLRSLDRVLCVLDAL 244
QY 233 PLDAQASHWPAFTCLALLCGFLPLALLMLCYGATLTLAASGRYGHALRLTAVLASA 292
|||
DB 245 PLTEGTSHWRPACILAVLGCVFPLAMGLCYGATLALAANGORYSHALRLTAVLASA 304
|||
QY 293 VAFFVPSNLLILLHYSDPSAMGNLYGAYVPSLALSTLNSCVDPFYIYVSAEFRKVR 352
|||
DB 305 VASFTPSNVLVHLVSNPSPEAGNLYGAYVPSLALSTLNSCVDPFYIYVSHFEKVR 364
|||
QY 353 AGLFORSFGDTVASKASAEGRSGMGTHSSLL 384
|||
DB 365 A-MLCQPEASSSSQASREAGSRGTAICSSLT 395
|||

RESULT 3
O93361 PRELIMINARY; PRT; 328 AA.
AC O93361;
DT 01-NOV-1998 (Tremblrel. 08, Created)
DT 01-NOV-1998 (Tremblrel. 08, Last sequence update)
DT 01-MAY-1999 (Tremblrel. 10, Last annotation update)
DE G PROTEIN-COUPLED RECEPTOR P2Y3.
GN P2Y3.
OS Meleagris gallopavo (Common turkey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Archosauria; Aves;
OC Neognathae; Galliformes; Meleagrididae; Meleagris.
RN [1]
RP SEQUENCE FROM N.A.
RA LI Q., OLESKY M., PALMER R.K., HARDEN T.K., NICHOLAS R.A.;
RT "Evidence that the p2y3 receptor is the avian homologue of the
RT mammalian p2y6 receptor.";
RL Mol. Pharmacol. 0:0-0(1998).
DR EMBL; AF069555; AAC23863.1; -.
DR PFAM; PF00001; 7tm1; 1.
SQ SEQUENCE 328 AA; 37594 MW; 1E0F2067 CRC32;
```

Query Match 17.9%; Score 357; DB 13; Length 328;
Best Local Similarity 28.7%; Pred. No. 1e-19;
Matches 89; Conservative 63; Mismatches 122; Indels 36; Gaps 10;

```
QY 79 LVPALYGLVLYVGLPANGLA---LWLATQAPRLPSTMLMLNLTADLLALPLPRIAY 135
|||
DB 23 LPLVYVVELLGLPLNAVWIGQIWLARKALTR--TTIYMLNLTADLLYCSLPILIYN 80
|||
QY 136 HLRGWPFGEAACLATAAALYGHMYGVLAAVSLDRYLALVHLPLA--RALRGRLAL 194
|||
DB 81 YTKDYWPFGDPTCKFVRFPYFTNLHGSILFTLCISVQRYMGICHPLASWHKKKKLTW 140
|||
QY 195 GLCMAAWLMAALALPLTLQRTFLA---RSDRVLCHDALPLDAQASHWPAFTCLAL 250
|||
DB 141 LYCAAVWIVTAQCLP-----TFVFASTGTQRTNVCYDLSPPDRSASYPPYGIT-LTI 193
|||
QY 251 LGCFLPLALLMLCYGATLTLAASGRYGHALR-----LTAVLASAVAF--VP 298
|||
DB 194 TGFLLPFAAILACYCSMARILCKQDELIGLAVHKKKAVRMIIVIVFSISFPFHLT 253
|||
QY 299 SNLILLHYSDPSAMGNLYG-AYVPSLALSTLNSCVDPFYIYVSAEFRKVRAGLFQ 357
|||
DB 254 KTIYLVIRSSPTLPCTLOAFAIAYKCTRPFASMNVLDPILFYFTQKFKRESTRYLL-- 312
|||
QY 358 RSPGDTVASK 367
|||
DB 312 ----DKMSK 317
|||
```

RESULT 4
O57466 PRELIMINARY; PRT; 374 AA.
ID O57466
AC O57466;
DT 01-JUN-1998 (Tremblrel. 06, Created)
DT 01-JUN-1998 (Tremblrel. 06, Last sequence update)

DT 01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
 DE G PROTEIN COUPLED P2Y NUCLEOTIDE RECEPTOR.
 GN TPZY.
 OS Meleagris gallopavo (Common turkey).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Archosauria; Aves;
 OC Neognathae; Galliformes; Meleagridae; Meleagris.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=BLOOD;
 RX MEDLINE; 98086419.
 RA BOYER J.L., WALDO G.L., HARDEN T.K.;
 RT "Molecular cloning and expression of an avian G protein-coupled P2Y
 receptor."
 RL Mol. Pharmacol. 52:928-934(1997).
 DR EMBL; AF031897; AAC60339.1; -;
 DR PFM; PF00001; 7tm1; 1.
 SQ SEQUENCE 374 AA; 42594 MW; 5320428C CRC32;

Query Match 17.7%; Score 353; DB 13; Length 374;
 Best Local Similarity 27.0%; Pred. No. 2.3e-19;
 Matches 89; Conservative 67; Mismatches 140; Indels 34; Gaps 9;

QY 79 LVPALYGLVVLVGLPANGALWLVATO-APRLPSTMLMNLATADLLALALPPRIAYHL 137
 DB 43 LLLPSIGIVFVGLPLNSWANNIFVSEMRPNWNTTYNFENLAISDTLYVESLTLVYYA 102

QY 138 RQGRWPFGEAACRLATAALYGHMYGVSLLAAVSLDRYLALVHLRRLARALGRRLALGLC 197
 DB 103 DRNNWPGKVFCKIVRFLEFVANLYSSILFTICISVHRMGICHPIRSKWKVKHARLIC 162

QY 198 MAWLMAALALPLTLQRTFLRSRDRVCHDALPDAQASHWQAFCTCLALGGLPL 257
 DB 163 VGVNLVVTICLIPNLIFVTT--SKDNSTLCHDTTRPE-EFDHYHYSSINALLFGIPF 219

QY 258 LAMLCYGAHLTL------AASGRYGH-ALRLTAVVLASAVAFVPSNLLLLHY 308
 DB 220 LVIVCYCLMAKRLCKRSFSPRVPYSYKRSIKMIIIVTFAICFVFHTRILYIT 279

QY 309 -----DPSAWGNYGAYVPSLALSTLNSCVDPFIYVSAEFDKVRAGLQF----- 359
 DB 280 SRYFOADQCLNINFTYKITRPLASINSLDPIYFMAGDKYGRRLRGAQRPRVPT 339

QY 359 -----SP-----GTVASKASAEGRSGMT 379
 DB 340 SLLALVSPSDSVSGCCNCE--SRGSGMT 367

RESULT 5
 ID O15132 PRELIMINARY; PRT; 370 AA.
 AC O15132;
 DT 01-JAN-1998 (TrEMBLrel. 05, Created)
 DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
 DT 01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
 DE P2Y5-LIKE RECEPTOR.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 97366605.
 RA JANSSENS R., BOEYNAEMS J.M., GODART M., COMMUNI D.;
 RT "Cloning of a human heptahelical receptor closely related to the P2Y5
 receptor."
 RL Biochem. Biophys. Res. Commun. 236:106-112(1997).
 DR EMBL; AF005419; AAB66322.1; -;
 DR PFM; PF00001; 7tm1; 1.
 SQ SEQUENCE 370 AA; 41861 MW; A42AB01D CRC32;

Query Match 17.4%; Score 345.5; DB 4; Length 370;
 Best Local Similarity 29.2%; Pred. No. 8.5e-19;

Matches 95; Conservative 61; Mismatches 138; Indels 31; Gaps 9;

QY 37 DSTPSILPAPRGYPGQVCANDSTLELPSSRALLGLWVPTRLVPALYGLVLVGLPANG 96
 DB 13 DSNSSLRPR-----LGNATAN--NTCIVDSDSKYKLNGLNG-----AVYSVFTLGLITNS 58

QY 97 LALWVLATQAP-RLPSTMLMNLATADLLALALPPRIAYHLRQGRWPFGEAACRLATAA 155
 DB 59 VSLFVFCFRMKMRSETAIFITNLAVSDLLFVCTLPKIFYNF-NRHWPFGDTLCKISGTA 117

QY 156 LYGHMYGVSLLAAVSLDRYLALVHLRRLARALGRRLALGLCMAALMAALPLTLQRT 215
 DB 118 FLTNIYGSMLFTCISVDRELAIVYFPRSTIRTRNSAIVCAGWILVLSGGISASLFS 177

QY 216 QTFRLARSDRVLCCHDALPDAQASHWQAFCTCLALGGLPLMLLCLCYGATLH----- 270
 DB 178 TT--NVNNATTTCFEGLSKRVKTYLSKITFIEVVGFIITLNLVSCSVSVLRLTKPA 235

QY 270 TLAASGRYGHALRLTAVVLASAVAFVPSNLLLLHYSDPSAWGNLY-----GAYV 323
 DB 236 TISQIGTNKKVKLMTIVHMAVFCVFPYNSVFLF-YALVRSQAITNCFLEFAKIMYP 294

QY 324 PSLALSTLNSCVDPFIYVSAEFR 348
 DB 295 IITCLATLNCCDPFIYVTSLEFQ 319

RESULT 6
 ID O97666 PRELIMINARY; PRT; 380 AA.
 AC O97666;
 DT 01-MAY-1999 (TrEMBLrel. 10, Created)
 DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
 DT 01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
 DE PUTATIVE CHEMOKINE RECEPTOR.
 GN APJ.
 OS Macaca mulatta (Rhesus macaque).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Primates; Catarrhini; Cercopitheidae; Cercopitheinae;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=SPLEEN;
 RA MARGULIES B.J., HAUER D.A., CLEMENTS J.E.;
 RT "Identification and characterization of thirteen rhesus macaque
 chemokine receptors and chemokine receptor homologues."
 RL Submitted (Oct-1998) to the EMBL/GenBank/DBJ databases.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
 DR EMBL; AF100206; AAC7404.1; -;
 DR PROSITE; PS00237; G-PROTEIN RECEPTOR; 1.
 KW Receptor; G-protein coupled receptor; Transmembrane; Glycoprotein.
 SQ SEQUENCE 380 AA; 42606 MW; D209A573 CRC32;

Query Match 17.3%; Score 344; DB 6; Length 380;
 Best Local Similarity 29.5%; Pred. No. 1.1e-18;
 Matches 97; Conservative 59; Mismatches 141; Indels 32; Gaps 9;

QY 79 LVPALYGLVVLVGLPANGALWLV--ATQAPRLPSTMLMNLATADLLALALPPRIAYH 136
 DB 30 LIPAIYMLVFLLTGTGNGLVITVFRSSKRKSADIFTASLAVALQTFVVTIPLWATYT 89

QY 137 LRQGRWPFGEAACRLATAALYGHMYGVSLLAAVSLDRYLALVHLRRLARALGRRLALGL 196
 DB 90 YRDYDWPFGFTFCKLSYLLFVNMYASVCLTGLSDRLAIIVRPVANRLRVSQAVA 149

QY 197 CMAALMAALALPLTLQRTFLRSRDRVLCCHDALPDAQASH-----WQPAF-TCIALLG 252
 DB 150 TAVLWVLAAALLAMPVNVFRTTGDLNTTKVQCYMDYSWATVSSDVAWEVGLGVSTTVG 209

QY 253 CFLPLAMLCYCATLHTLAASGR-----YGHALRLTAVVLASAVAF---FVPSNLLLL 304

Db 210 FVPPTIMLTCTYFFIAQTAGHFRKERIEGLRKRRLLSIIIVLVVTFALCWMPYHLVKT 269
Qy 305 LHYSPPSPSAMNLYGAVPSL-----ALSTLNSCVDPFIYIVVSAEPRDKVR----- 353
Db 270 L-YMLGSLHLWPCDFDLFMVFPYCTCISYVNSCLNFILYAFFDPRFQACTSMLCGG 328
Qy 353 ---AGLFORSFGDTVASKASAEGRSGRMG 378
Db 329 SRCATSHSSSGEKSASYSS--GHSQGP 355
RESULT 7
ID 070129 PRELIMINARY; PRT; 345 AA.
AC 070129;
DT 01-AUG-1998 (Tremblrel. 07, Created)
DT 01-AUG-1998 (Tremblrel. 07, Last sequence update)
DE 01-MAY-1999 (Tremblrel. 10, Last annotation update)
DE ANAPHYLATOXIN C5A RECEPTOR.
OS Cavia porcellus (Guinea pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Hystricognathi; Caviidae; Cavia.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-HARTLEY;
RX MEDLINE; 98236152.
RA FUKUOKA Y., EMBER J.A., YASUI A., HUGLI T.E.;
RT "Cloning and characterization of the guinea pig C5a anaphylatoxin
RT receptor: interspecies diversity among the C5a receptors.";
RL Int. Immunol. 10:275-283(1998).
DR EMBL; U86103; AAC40074.1; -.
DR PFAM; PF00001; 7tm_1; 1.
SQ SEQUENCE 345 AA; 38592 MW; A47BBFC0 CRC32;

Query Match 17.1%; Score 340.5; DB 11; Length 345;
Best Local Similarity 27.3%; Pred. No. 1.9e-18;
Matches 94; Conservative 62; Mismatches 149; Indels 39; Gaps 6;

Qy 57 DSDTLELPDSSRALLGWPTRLVPALYG-----LVLVVGPLPANGALWLATQAP 107
Db 10 DYNSTFLDP-----GFVDNVERLSFGDLVAVVMVFLVGVPGNALVWVTACEAR 62
Qy 108 RLPSTMLMNLNATADLLALALPPRIAYHLRQGRWPFGEAACRLATAALYGHMYSVLL 167
Db 63 RHINAIWFLNLAADLLSCLAPILLVSTVHLNHWYFGDTACKVLPSSLINMYTSILL 122
Qy 168 AAVSLDRYALVHLPLRALRGRRLALGLCHNAWLMALALPLTLQRTFLARSDRYL 227
Db 123 ATISADRLULLVSLPWCQRFRGGCLAWTACGLAWYALLLSPPSLYRRTTNEHFSFKY 182
Qy 228 CHDALPLDAQASHWQPAFTCLALLGCLFLLPMLLCYG-ATLHTLAASGRYGHALRLTA 286
Db 183 C--VTDYGRDISKERAVALVLLVGFVPLTLTACYTLFLLRTWSRKATRSKTVKVV 240
Qy 287 VVLASAVAFFVPSNLL-LLHYSDPSPSANGNLGYAVPSLSTLNSCVDPFIYIVVSA 345
Db 241 AVVSSFFVFWLPYQVTGILLAWHSPNSATYRNTKALDVCVAFAYINCCINPIIVVWAG 300
Qy 346 EFRDKVRAGL-----FQSPGDTVASKASA 370
Db 301 GFQGRLLKSLPSVLRNVLTESLDRKHOSFARSTVDTMPQKSES 344

RESULT 8
ID Q92280 PRELIMINARY; PRT; 371 AA.
AC Q92280;
DT 01-MAY-1999 (Tremblrel. 10, Created)
DT 01-MAY-1999 (Tremblrel. 10, Last sequence update)
DE 01-MAY-1999 (Tremblrel. 10, Last annotation update)
DE GALANIN RECEPTOR 2.
OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-129/SV;
RX MEDLINE; 99048962.
RA PANG L., HASHEMI T., LEE H.J., MAGUIRE M., GRAZIANO M.P., BAYNE M.,
RA HAWES B., WONG G., WANG S.;
RT "The mouse GalR2 galanin receptor: genomic organization, cDNA
RT cloning, and functional characterization.";
RL J. Neurochem. 71:2252-2259(1998).
DR EMBL; AF077375; AAC95468.1; -.
KW Receptor.
SQ SEQUENCE 371 AA; 40883 MW; 4B6912E5 CRC32;

Query Match 16.9%; Score 336; DB 11; Length 371;
Best Local Similarity 32.5%; Pred. No. 4.4e-18;
Matches 111; Conservative 44; Mismatches 157; Indels 30; Gaps 11;

Qy 56 NDSPTLELPDSSRALLGWPTRL-LVPALYGLVAVGLPANGALWLATQAPRLPST-M 113
Db 2 NGSDSQGAEDSSQGGGQWQPEAVLPLFFALIFLVGAVGNALVLVLRGGQAVSTTNL 61
Qy 114 LLMNLATADLLALALPPRIAYHLRQGRWPFGEAACRLATAALYGHMYSVLLAAVSLD 173
Db 62 FILNLGVADLCFILCCVPFQATITLDDWFGSLCKAVHFLIFTMHASSFTLAAVSLD 121
Qy 174 RYLAIVHLPLRALRGRRLALGLCHMAWLMALALPLTLQRTFLARSDRYLCHDALP 233
Db 122 RYLAIRYPLSHRELTPRNALAAIGLIWGLALLFGPYLSYSQSOLA--NLTVCHPA-- 178
Qy 234 LDAQASHWQPAFTCLALLGCLFLLPMLLCYGATLHTL-----AASGRYGHALRLT 285
Db 178 WSAPRRFNSCTFCLSYL---LPVLVLSLYARTLTLWRTVDPVAVAGSSQRAKR-KVT 233
Qy 286 AVVLASAVAF---FVPSNLLLLHYSDPSPSANGNLGYAVPSLSTLNSCVDPFIYIVV 342
Db 234 RMIVAVVFLCWLMPHALLILCVWFGFRLPTRLAT-YALRILSHLVSVANSVNPVIVAL 292
Qy 343 VSAEFRD---KVRAGLFQSPGDTVASKASAEGRSGRMGTHS 381
Db 293 VSKHFRKGRKICAGLLRRAP-----RRASGRVCILAXGNHS 329

RESULT 9
ID O35811 PRELIMINARY; PRT; 361 AA.
AC O35811;
DT 01-JAN-1998 (Tremblrel. 05, Created)
DT 01-JAN-1998 (Tremblrel. 05, Last sequence update)
DT 01-NOV-1998 (Tremblrel. 08, Last annotation update)
DE G-PROTEIN COUPLED RECEPTOR.
GN P214.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-SPRAGUE-DAWLEY; TISSUE=LIVER;
RA BOGDANOV Y.D., WILDMAN S., KING B.F., BURNTOCK G.;
RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-SPRAGUE-DAWLEY; TISSUE=BRAIN;
RX MEDLINE; 98421785.
RA WEBB T.E., HENDERSON D., ROBERTS J.A., BARNARD E.A.;
RT "Molecular cloning and characterization of the rat p2y4 receptor.";
RL J. Neurochem. 71:1424-1434(1998).
DR EMBL; Y14705; CAA75007.1; -.
DR EMBL; Y11433; CAA72241.1; -.
DR PFAM; PF00001; 7tm_1; 1.
KW G-protein coupled receptor.

SQ SEQUENCE 361 AA; 40893 MW; D1882F0C CRC32;

Query Match 16.7%; Score 332; DB 11; Length 361;

Best Local Similarity 29.5%; Pred. No. 8.5e-18;

Matches 94; Conservative 48; Mismatches 157; Indels 20; Gaps

6;

QY 79 LVPALYGLVLPVGLPANGALWV-LATOAPRLPSTMLLMNLTADLLALALPPIAYHL 137

DB 33 LLPSYAVVFLGLNALNAPTLEFLRPMWDATATYFHLALSOTLYVLSLPTLVYYA 92

QY 138 RGQWPGEACRATATYALYCHMYGVSLLAAVSLDRYLALVHPLRPARLGRRLALGLC 197

DB 93 ARNHPEFTGLCKFVRFYFNWLVCSVLFCTCISVHYRGLGICHPRAIRWGRPRFASLLC 152

QY 198 MAALMAALALPTLOROTFLARSDVLCHDALPDAQASHQWPAFTCLALGCFPL 257

DB 153 LGVLLVAGCLVPLNFFVTT--NANGTILCHDT-TLPEEDHYVPSSAVVLLGLPF 209

QY 258 LAMLCYGATLHTL-----AASRRYGHARLTAVVLASAVAFVPSNLLLLHSDPS 311

DB 210 LITLVCYGLMARLYRPLPGAGQSSRLSRTIATVLTFAVCFVPPHITRTIYYQARL 269

QY 312 PSANGLYG-----AYPSLALSTLNSCVDPEIYVVSNAEFDKVRAGLFQSPGDTVASK 367

DB 270 LOADCHLVINNVYKVRPLRPLANSCLDPVLVLTGDKYRNQLOQLCGRSKPKRPTAAS 329

QY 368 ASA-----EGSGRGMTH 380

DB 330 SLALVTLHEESISRWDTH 348

RESULT 10

ID O88854 PRELIMINARY; PRT; 371 AA.

AC O88854;

DT 01-NOV-1998 (TREMBlrel. 08, Created)

DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)

DT 01-MAY-1999 (TREMBlrel. 10, Last annotation update)

DE GALANIN RECEPTOR TYPE 2.

GN GALR2.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-129/SV;

RA KOLAKOWSKI L.F. JR., BROUSSARD S.R.;

RT "Galanin Receptor Type 2 (GALR2) from Human and Mouse: Genomic

RT Cloning, Chromosomal Localization, Functional Expression, Gq

RT Dependent Signal Transduction, and Expression Pattern.";

RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF042784; AAC36589.1; -

DR PFAM; PF00001; 7tm_1; 1.

SQ SEQUENCE 371 AA; 40533 MW; FF7FBA0C CRC32;

Query Match

Best Local Similarity 16.1%; Score 321.5; DB 11; Length 371;

Matches 107; Conservative 46; Mismatches 160; Indels 29; Gaps

10;

QY 56 NDSOTLEPDSRRALLGWVTR-LVPALYGLVLPVGLPANGALWVLTATQAPRLPST-M 113

DB 2 NGDSQGAEDSQGGGQWQPEAVLPLFFALIFLVGAVGNALVAVLLRGQAVSTNL 61

QY 114 LLMNLTADLLALALPPIAYHLRGQWPGEACRATATYALYCHMYGVSLLAAVSLD 173

DB 62 FILNLGADLCFLCCVFPQATITVLDQWVFGSLCKAVHFLITMTHASSTFLAAVSLD 121

QY 174 RYLAHVPLRARRGRRLRGLCLGCMMAALPLTLQRTQRLRSLDRVLCHDALP 233

DB 122 RYLAIRPMHRELTPNALAAGLWGLALLFGPILSYSSQLA--NLTVCHPAW- 179

QY 234 LDAQASHQWPAFTCLALGCFPLPPLAMLCYCATLHTL-----AASRRYGHARLT 285

DB 179 ---SAPRRRAMDLCTFVSFYLPLVLSLTARTLHYLWRTVDFVAGSGSORAKR-KVT 234

QY 286 AVVLASAVF---FVPSNLLLLHSDPSANGNLYGAVVPSLALSTLNSCVDPEIYY 342

DB 235 RMIVIVAVLFCWMPHMHALLCVMWFORFPLTRAT-YALRILSHLVSYANSCVPIVIAL 293

QY 343 VSAEFD---KVRAGLFQSPGDTVASKAEGSGRGMTHS 381

DB 294 VSRHFRGFKICAGLLRRAP-----RRASGRVCILAPGNHS 330

RESULT 11

ID O43603

AC O43603 PRELIMINARY; PRT; 387 AA.

DT 01-JUN-1998 (TREMBlrel. 06, Created)

DT 01-JUN-1998 (TREMBlrel. 06, Last sequence update)

DT 01-MAY-1999 (TREMBlrel. 10, Last annotation update)

DE GALANIN RECEPTOR GALR2.

GN GALR2 OR GALR2.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

OC Eutheria; Primates; Catarrhini; Hominoidea; Homo.

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE; 98153789.

RA BLOOMQUIST B.T., BEAUCHAMP M.R., ZHELNIN L., BROWN S.-E.,

RA CORE-WILLSE A.R., GREGOR P., CORNFELD L.J.;

RA "Cloning and expression of the human galanin receptor Galr2.";

RL Biochem. Biophys. Res. Commun. 243:474-479(1998).

RN [2]

RP SEQUENCE FROM N.A.

RA FATHI Z., BATTAGLINO P.M., IBEN L.G., LI H., BAKER E., ZHANG D.,

RA MCGOVERN R., MAHLE C.D., SUTHERLAND G.R., IISMAA T.P.,

RA DICKINSON K.E.J., ANTAL ZIMANYI I.;

RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.

RN [3]

RP SEQUENCE FROM N.A.

RA KOLAKOWSKI L.F. JR., BROUSSARD S.R.;

RT "Galanin Receptor Type 2 (GALR2) from Human and Mouse: Genomic

RT Cloning, Chromosomal Localization, Functional Expression, Gq

RT Dependent Signal Transduction, and Expression Pattern.";

RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.

RN [4]

RP SEQUENCE FROM N.A.

RX MEDLINE; 99094693.

RA BORORSKY B., WALKER M.W., HUANG L.Y., JONES K.A., SMITH K.E., BARD J.,

RA BRANCHER T.A., GERALD C.;

RT "Cloning and characterization of the human galanin GALR2 receptor.";

RL Peptides 19:1771-1781(1998).

DR EMBL; AF040630; AAC39634.1; -

DR EMBL; AF058762; AAC18118.1; -

DR EMBL; AF042782; AAC36587.1; -

DR EMBL; AF080386; AAD08671.1; -

DR PFAM; PF00001; 7tm_1; 1.

KW Receptor.

SQ SEQUENCE 387 AA; 41700 MW; E7A343C3 CRC32;

Query Match

Best Local Similarity 16.1%; Score 320; DB 4; Length 387;

Matches 111; Conservative 46; Mismatches 140; Indels 36; Gaps

15;

QY 73 GWVTR-LVPALYGLVLPVGLPANGALWVLTATQAPRLPST-MLLMNLTADLLALALP 130

DB 20 GWHPEAVIVPLLFALIFLVGTVGNTLVAVLLRGQAVSTNLFILNLGVADLCFLCCV 79

QY 131 PRIA--YHLRGQWPGEACRATATYALYCHMYGVSLLAAVSLDRYLALVHPLRARRAL 188

DB 80 PFQATITLQD--WVFGSLCKAVHFLITMTHASSTFLAAVSLDRYLALVHPLRARRAL 137

QY 189 GRLALGLCMAAWLMAALALP-LTQOTFLRLASDRVLCHDALPLDAQASHWQPAFTC 247
| | | | | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 138 TPNALAAIGLWGLSULSGPVSRYROS-QLA--NLIVCHPAW-----SAPRRAMDIC 190
| | | | | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 248 LALLGCFPLPMLLCYCATLHTL-----AASGRYGHALRLPVLAVASAVAF---F 296
| | | | | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 191 TTFVSLLPLVLGLVLYARTLRLVRAVDPVAAAGSGARRAKR-KVTRMILVLAALPCLCW 249
| | | | | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 297 VPSNLLLLHSDPPSPSANGNLYGAVPSLSTLNSCVDPPFIYYVYSAEFRDKVR---A 353
| | | | | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 250 MPHAILCWFQGFPTLRAT-YALRLSLVSYANSCVPIYALVSKHFRKGFRTICA 308
| | | | | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 354 GLFQSPGDTVASKAASGSGRMGTHS-SLLQ 385
| | | | | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 309 GLLGRAPG-----RASGRVCAAAARGTHSGSVLE 336
| | | | | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
RESULT 12
O88626 PRELIMINARY; PRT; 370 AA.
AC O88626;
DT 01-NOV-1998 (TREMREL. 08, Created)
DT 01-NOV-1998 (TREMREL. 08, Last sequence update)
DE GALANIN RECEPTOR TYPE 3.
GN GALR3.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
[1]
RP SEQUENCE FROM N.A.
RC STRAIN-SPRAGUE DAWLEY; TISSUE-HYPOTHALAMUS;
RA WATERS S.M., KRAUSE J.E.;
RT "Rat galanin receptor type 3 (GALR3) coding region.";
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF079844; AAC34590.1; -;
DR PFAM; PF00001; 7tm_1; 1.
SQ SEQUENCE 370 AA; 40410 MW; 1EBF1E77 CRC32;

Query Match 15.9%; Score 317.5; DB 11; Length 370;
Best Local Similarity 33.7%; Pred. No. 1.1e-16;
Matches 104; Conservative 32; Mismatches 126; Indels 47; Gaps 10;

QY 73 GWPTRLVPLXGLVVLGVLGALWVL-----ATQAPRLPSTMLMLNLATADLLA 126
| | | | | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 13 GSVGAVAVPFIAPLIFLLGVNGVLAVLLOPGPSAWQEPSRTTDLFILNLAVADLCFI 72
| | | | | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 127 LALPPRIAYHLRGQRPFGAECRLATAALYGHMYSVLLLAASVLDRLYALVHPLRARA 186
| | | | | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 73 LCCVPFQAAIYTLDAWLFQAFVCKTVHLLIYLTMTWASFTLAASVLDRLYAVRHPLRSRA 132
| | | | | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 187 L--RGRRLALGLCMAAWLMAALALP-LTQOTFLRLASDRVLCHDALPLDAQASHWQ 243
| | | | | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 133 LATPNRAAVGL---VWLLALFSAPIYLSYGTGTVR-----YGALELCVPA--WED 178
| | | | | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 244 A-----FTCLALLGCFPLPMLLCYCATLHTL-----AASGRR--YCHALRLTA 286
| | | | | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 179 ARRRALDVATFAAGYLLPVAVVSLAYGRTLCFLWAAGVPAGAAAEARRATGRAGAML 238
| | | | | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 287 VVLASAVAFVPSNLLLLHSDP-----SPSANGNLYGAVPSLSTLNSCVDPPFIYYV 343
| | | | | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 239 AVAALYALCWGPHHALILCFWYGRFAFSPAT-----YACRLASHCLAYANSCNLPVYSLA 294
| | | | | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 344 SAEFRDKVR 352
| | | | | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 295 SRHFRARFR 303
| | | | | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
RESULT 13
O93239 PRELIMINARY; PRT; 342 AA.
ID O93239
AC O93239;

DT 01-NOV-1998 (TREMREL. 08, Created)
DT 01-NOV-1998 (TREMREL. 08, Last sequence update)
DT 01-MAY-1999 (TREMREL. 10, Last annotation update)
DE CXK CHEMOKINE RECEPTOR-2.
OS Cyprinus carpio (Common carp).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Actinopterygii;
OC Neopterygii; Teleostei; Euteleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Cyprinidae; Cyprinidae; Cyprinus.
[1]
RP SEQUENCE FROM N.A.
RA FUJIKI K., NAKAO M., SHIN D., YANO T.;
RT "cDNA cloning of a carp homologue-2 of mammalian interleukin-8
receptors.";
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB010713; BAA31470.1; -;
DR PFAM; PF00001; 7tm_1; 1.
SQ SEQUENCE 342 AA; 38481 MW; D3370D68 CRC32;

Query Match 15.8%; Score 314; DB 13; Length 342;
Best Local Similarity 30.7%; Pred. No. 1.8e-16;
Matches 99; Conservative 53; Mismatches 129; Indels 42; Gaps 13;
QY 85 GLVLY-----VGLPANGALWVLATQAPRLPST-MLLMLNLATADLLALPPRIAYHL 137
| | | | | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 28 GLVLYVSVFCLSLPGNMVIVFVWSCMENRRSTDVILMHLAIADLFFALTLPFSAA-DI 86
| | | | | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 138 RQRPFGAECRLATAALYGHMYSVLLLAASVLDRLYALVHPLRARRLRALGIC 197
| | | | | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 87 HAGHWMFGTFMCKLISGLQEATFYCCVFLACISVDRLAIVKATQFLAQK-RHLVGIVC 145
| | | | | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 198 MAALWMAALALP-LTQOTFLRLASDRVLCHDALPLDAQASHWQPAFTCL-ALLGCFLP 256
| | | | | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 146 ALVWVCAFLLSLPVIVNQAFITGNMGDIYCHDNVATSGMDS-WRMGLRIHLHTLGFELP 204
| | | | | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 257 LLAALLCYCATLHTLAASGRYGHALRLTAVVLASAVAFV---PSNLLLL----- 306
| | | | | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 205 LAVMMFCYGTMTCTCRT--RNSQKQKAMRVLSVVLAFIVCWLFPNILELIDITLMRGK 262
| | | | | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 306 -----HYSDPSPSANGNLYGAVPSLSTLNSCVDPPFIYYVSAEFRDKVRAGLFORS- 360
| | | | | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 263 VIETCELKDSIDVA-----LYV-TQMAFAHCAINPLIYAFIGKFRNQLMLSLFKRGL 315
| | | | | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 360 PGDTVASK-----ASAEGSGRM 377
| | | | | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 316 LGRNTMSKYRVGSVNSTGSTRQM 338
| | | | | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |

RESULT 14
O60755 PRELIMINARY; PRT; 368 AA.
AC O60755;
DT 01-AUG-1998 (TREMREL. 07, Created)
DT 01-AUG-1998 (TREMREL. 07, Last sequence update)
DT 01-MAY-1999 (TREMREL. 10, Last annotation update)
DE GALANIN 2 RECEPTOR HOMOLOG.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominiidae; Homo.
[1]
RP SEQUENCE FROM N.A.
RC TISSUE-UTERUS;
RA BENNETT M.M., LESCOE M.K., GALLIPOLI P.2., RAMABHADRAN T.V.;
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
[2]
RP SEQUENCE FROM N.A.
RX MEDLINE; 98389766.
RA SMITH K.E., WALKER M.W., ARTYMYSHYN R., BARD J., BOROWSKY B.,
RA TAMM J.A., YAO W.-J., VAYSSE P.J.-J., BRANCHEK T.A., GERALD C.,
RA JONES K.A.;
RT "Cloned human and rat galanin GALR3 receptors. Pharmacology and
activation of G-protein inwardly rectifying K+ channels.";
J. Biol. Chem. 273:23321-23326(1998).

This Page Blank (uspt)

GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 28, 1999, 19:54:39 ; Search time 1202.25 Seconds
(without alignments)
8031.234 Million cell updates/sec

Title: US-09-371-333-1
Perfect score: 4895
Sequence: 1 CTCGCCAGCGCTGGCTGGCA.....AAAAAAAAAAAAAAAAAAAA 4895

Scoring table: IDENTITY_NUC
Searched: 2546578 seqs, 986266752 residues

Database : EST.*

- 1: em_est1.*
- 2: em_est2.*
- 3: em_est3.*
- 4: em_est4.*
- 5: em_est5.*
- 6: em_est6.*
- 7: em_est7.*
- 8: em_est8.*
- 9: em_est9.*
- 10: em_est10.*
- 11: em_est11.*
- 12: em_est12.*
- 13: em_est13.*
- 14: em_est14.*
- 15: em_est15.*
- 16: em_est16.*
- 17: em_est17.*
- 18: em_est18.*
- 19: em_est19.*
- 20: gb_est1.*
- 21: gb_est2.*
- 22: gb_est3.*
- 23: gb_est4.*
- 24: gb_est5.*
- 25: gb_est6.*
- 26: gb_est7.*
- 27: gb_est8.*
- 28: gb_est9.*
- 29: gb_est10.*
- 30: gb_est11.*
- 31: gb_est12.*
- 32: gb_est13.*
- 33: gb_est14.*
- 34: gb_est15.*
- 35: gb_est16.*
- 36: gb_est17.*
- 37: gb_est18.*
- 38: gb_est19.*
- 39: gb_est20.*
- 40: gb_est21.*
- 41: gb_est22.*
- 42: gb_est23.*
- 43: gb_est24.*
- 44: gb_est25.*
- 45: gb_est26.*
- 46: gb_est27.*
- 47: gb_est28.*
- 48: gb_est29.*
- 49: gb_est30.*
- 50: gb_est31.*
- 51: gb_est32.*
- 52: em_est20.*
- 53: em_est21.*

54: em_est22.*
55: em_est23.*
56: em_est24.*
57: em_est25.*
58: em_est26.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	553.4	11.3	570	51	AI739295	AI739295 w130b03.x
C 2	530	10.8	544	45	AI344017	AI344017 tc01e02.x
C 3	291.8	6.0	611	53	HS0005397	AI040921 Homo sapi
C 4	282	5.8	985	54	HS0009189	AI044339 Homo sapi
C 5	278.6	5.7	542	28	AA081138	AA081138 zn34h07.s
C 6	274	5.6	540	49	AL048626	AL048626 DKF2p564F
C 7	261.2	5.3	529	37	AA680243	AA680243 ac86a01.s
C 8	259.6	5.3	504	35	AA577824	AA577824 nn24e06.s
C 9	259.2	5.3	487	37	AA708108	AA708108 zg06b11.s
C 10	257.2	5.3	457	31	AA284247	AA284247 zc55d01.T
C 11	255.2	5.2	496	41	AI038990	AI038990 ox24f02.x
C 12	255.2	5.2	487	50	AI679002	AI679002 tu60g01.x
C 13	253	5.2	490	49	AI653525	AI653525 tq95e03.x
C 14	252.8	5.2	523	41	AI054414	AI054414 q176h03.x
C 15	252	5.1	569	49	AL048969	AL048969 DKF2p434M
C 16	251.6	5.1	521	44	AI302688	AI302688 qn17a10.x
C 17	247.6	5.1	500	51	AI732911	AI732911 oet2lg10.x
C 18	247.2	5.1	535	48	AI608771	AI608771 tw94f06.x
C 19	240.8	4.9	660	49	AL047602	AL047602 DKF2p586G
C 20	239.6	4.9	630	46	AI457389	AI457389 tl73c12.x
C 21	239	4.9	494	34	AA486896	AA486896 ab17h07.s
C 22	235.6	4.8	711	54	HS0009190	AI044340 Homo sapi
C 23	233.6	4.8	524	21	T63408	T63408 yc23e07.s1
C 24	233.6	4.8	443	47	AI468006	AI468006 tj84f09.x
C 25	231.6	4.7	552	35	AA594229	AA594229 nn29h05.s
C 26	231.2	4.7	543	36	AA608741	AA608741 ae56g03.s
C 27	230.8	4.7	441	51	AI732120	AI732120 ac86a01.x
C 28	230.8	4.7	441	51	AI732180	AI732180 ac85g01.x
C 29	230.8	4.7	646	54	HS0008595	AI043745 Homo sapi
C 30	230.8	4.7	591	54	HS0011470	AI046620 Homo sapi
C 31	230.2	4.7	732	53	HS0007755	AI042905 Homo sapi
C 32	229.8	4.7	474	45	AI354847	AI354847 qt76a04.x
C 33	229.4	4.7	430	53	HS0005489	AI041013 Homo sapi
C 34	229	4.7	563	36	AA601355	AA601355 no18b04.s
C 35	228.2	4.7	554	34	AA524604	AA524604 nh34c01.s
C 36	227.4	4.6	546	48	AI567106	AI567106 tq85c11.x
C 37	227	4.6	492	36	AA614180	AA614180 no94g11.s
C 38	227	4.6	783	54	HS0010804	AI045954 Homo sapi
C 39	226.4	4.6	440	36	AA618412	AA618412 nn27d01.s
C 40	226.4	4.6	478	50	AI683547	AI683547 tx67g09.x
C 41	226.2	4.6	435	34	AA487119	AA487119 ab19a06.s
C 42	226.2	4.6	474	34	AA493641	AA493641 nh03b10.s
C 43	226.2	4.6	492	34	AA534064	AA534064 nj97c05.s
C 44	226.2	4.6	604	35	AA573207	AA573207 nj42f12.s
C 45	225.6	4.6	455	40	AA904211	AA904211 od88e02.s

ALIGNMENTS

RESULT 1
AI739295/c
LOCUS AI739295 570 bp mRNA EST 18-JUN-1999
DEFINITION w130b03.x1 NCI_CGAP_Col6 Homo sapiens cDNA clone IMAGE:2391725 3'
similar to TR:076067 076067 PROTEASE-ACTIVATED RECEPTOR 4. ; mRNA
sequence.
ACCESSION AI739295
NID 95101276

BASE COUNT 98 a 182 c 203 g 61 t
ORIGIN Subtraction by Bento Soares and M. Fatima Bonaldo.

Query Match 10.8%; Score 530; DB 45; Length 544;
Best Local Similarity 99.3%; Pred. No. 6.3e-55;
Matches 544; Conservative 0; Mismatches 0; Indels 4; Gaps 1;
QY 245 AGCGTCTACGAGAGCGGAGCAGCGAGGTGGTGATGACACGCGCCCTCAATCCTG 304
DB 544 AGCGTCTACGAGAGCGGAGCAGCGAGGTGGTGATGACACGCGCCCTCAATCCTG 489
QY 305 CCGGCGCGCGCGCTACCGAGCGCAAGTCTGTGCCAATGACAGTACACCGCTGGAGCTC 364
DB 488 CCGGCGCGCGCGCTACCGAGCGCAAGTCTGTGCCAATGACAGTACACCGCTGGAGCTC 429
QY 365 CCGGAGAGCTACGCGGAGCTCTTCTGGGTGGGTGCCACAGGCTGGTGGCGCGCTC 424
DB 428 CCGGAGAGCTACGCGGAGCTCTTCTGGGTGGGTGCCACAGGCTGGTGGCGCGCTC 369
QY 425 TATGGGTGGTCTCTGGGTGGGTGCCGCAATGGGCTGGGCTGGTGGTGGTGGTGG 484
DB 368 TATGGGTGGTCTCTGGGTGGGTGCCGCAATGGGCTGGGCTGGTGGTGGTGGTGG 309
QY 485 ACGGAGGAGCTCGGCTGGGCTGCCGCGCAATGGGCTGGGCTGGTGGTGGTGGTGG 544
DB 308 ACGGAGGAGCTCGGCTGGGCTGCCGCGCAATGGGCTGGGCTGGTGGTGGTGGTGG 249
QY 545 CTGCTGGCGCTGGGCTGGGCTGCCGCGCAATGGGCTGGGCTGGTGGTGGTGGTGG 604
DB 248 CTGCTGGCGCTGGGCTGGGCTGCCGCGCAATGGGCTGGGCTGGTGGTGGTGGTGG 189
QY 605 TTCGGGAGGCGCGCTGGGCTGCCGCGCAATGGGCTGGGCTGGTGGTGGTGGTGGTGG 664
DB 188 TTCGGGAGGCGCGCTGGGCTGCCGCGCAATGGGCTGGGCTGGTGGTGGTGGTGGTGG 129
QY 665 GTGCTGTGCTGGCGCGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGG 724
DB 128 GTGCTGTGCTGGCGCGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGG 69
QY 725 GCCGCGCGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGG 784
DB 68 GCCGCGCGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGG 9
QY 785 GCGGCGCG 792
DB 8 GCGGCGCG 1

RESULT 3
HSM005397/c
ID HSM005397 standard; RNA; EST; 611 BP.
XX AL040921;
AC AL040921;
XX SV AL040921.1
XX NI e1399257
XX DT 12-MAR-1999 (Rel. 59, Created)
DT 12-MAR-1999 (Rel. 59, Last updated, Version 1)
XX Homo sapiens mRNA; EST DKFZp434J0515.s1 (from clone
DE DKFZp434J0515)
XX EST; expressed sequence tag.
KW
XX Homo sapiens (human)
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria;
OC Primates; Catarrhini; Hominidae; Homo.
XX
RN [1]

RP 1-611
RA Bloeker H., Boecker M., Brandt P., Mewes W., Gassenhuber J.,
RA Wiemann S.;
RT Submitted (12-MAR-1999) to the EMBL/GenBank/DBJ databases.
RL MIPS, Am Klopferspitz 18a D-82152 Martinsried, GERMANY
XX
CC Clone from S. Wiemann, sequenced by GBF within the cDNA
CC sequencing consortium of the German Genome Project
CC r1 sequence also available at the RZPD in Berlin
CC This clone is available at the RZPD in Berlin
CC Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
CC Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de
XX
FH Key Location/Qualifiers
FT source 1. .611
FT /db_xref="taxon:9606"
FT /organism="Homo sapiens"
FT /clone_lib="434 (synonym: htes3). Vector pSport1; host
FT DH10B; sites NotI + SalI"
FT /dev_stage="adult"
FT /tissue_type="testis"
XX
SQ Sequence 611 BP; 115 A; 168 C; 137 G; 191 T; 0 other;

Query Match 6.0%; Score 291.8; DB 53; Length 611;
Best Local Similarity 72.4%; Pred. No. 1.2e-26;
Matches 444; Conservative 0; Mismatches 142; Indels 27; Gaps 4;
QY 1561 GAGGCCAAGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1620
DB 608 GAGGCTGAGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 549
QY 1621 TAAACCCCATCTCTACCAAAATACAAATAGCTGGGCTGGTGGGCTGGGCTGGTGG 1680
DB 548 TGAACACTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 489
QY 1681 ATCCAGCTACTCAGGAGCTGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1740
DB 488 ATCCAGCTACTTGGGAGGCTGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 429
QY 1741 CAGTGGCCGAGATGCGCCACTGAGCTCCAGCTCCAGCTCCAGCTCCAGCTCCAGCTCC 1795
DB 428 CAGTGGCAGAGATCGCGGCTGAGCTCCAGCTCCAGCTCCAGCTCCAGCTCCAGCTCC 369
QY 1796 CTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1848
DB 368 CAAAAAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 309
QY 1848 -----CGGCGAGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCG 1898
DB 308 AGTGCAGTGTGCGGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTG 249
QY 1899 GATGGAGATGCTTGAAGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1958
DB 248 GAGCGGTGGAT-CACAAAGTTCAGGAGGTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 190
QY 1959 ATCTCTACACAAAAAATTTTATGAAACAGCAGGATTTGTCATGCGGCTATAGTCC 2018
DB 189 GTCTCTACTAAAAATACAA-----AAATTAAGCCAGGCGGTGGTGGGCGGCTGAGTCC 134
QY 2019 CAGCCACTCAAGAGCAGAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2078
DB 133 CAGCTACTCAGGCGGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 74
QY 2079 GAGCTATGATGTACCACTCCAGCTCCAGCTCCAGCTCCAGCTCCAGCTCCAGCTCCAGCT 2138
DB 73 GAGCCAAAGATCGTGCCACTCCAGCTCCAGCTCCAGCTCCAGCTCCAGCTCCAGCTCCAG 14
QY 2139 ATAACAAACTAA 2151

```
Db 13 AAAAAAAAAAAAA 1
|||||
RESULT 4
ID HSM009189
XX HSM009189 standard; RNA; EST; 985 BP.
AC AL044339;
XX AL044339.1
SV AL044339.1
XX e1402776
NI
XX 12-MAR-1999 (Rel. 59, Created)
XX 12-MAR-1999 (Rel. 59, Last updated, Version 1)
XX Homo sapiens mRNA; EST DKF2p434A092_r1 (from clone
DE DKF2p434A092)
XX EST; expressed sequence tag.
XX Homo sapiens (human)
XX Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria;
XX Primates; Catarrhini; Hominidae; Homo.
XX [1]
XX 1-985
XX Ansorge W., Wirkner U., Meves W., Gassenhuber J., Wiemann S.;
XX Submitted (12-MAR-1999) to the EMBL/GenBank/DBJ databases.
XX MIPS, Am Klopferspitz 18a D-82152 Martinsried, GERMANY
XX Clone from S. Wiemann, sequenced by EMBL within the cDNA
XX sequencing consortium of the German Genome Project
XX s1 sequence also available
XX This clone is available at the R2PD in Berlin
XX Please contact the R2PD: Ressourcenzentrum, Heubnerweg 6, 14059
XX Berlin-Charlottenburg, GERMANY; Email: clone@rpd.de
XX Key Location/Qualifiers
XX source
XX 1. .985
XX /db_xref="taxon:9606"
XX /organism="Homo sapiens"
XX /clone="DKF2p434A092"
XX /clone_lib="434 (synonym: htes3). Vector pSport1; host
XX DH10B; sites NotI + SalI"
XX /dev_stage="adult"
XX /tissue_type="testis"
XX
XX Sequence 985 BP; 334 A; 218 C; 234 G; 199 T; 0 other;
XX
Query Match 5.8%; Score 282; DB 54; Length 985;
Best Local Similarity 71.8%; Pred. No. 1.5e-25;
Matches 432; Conservative 0; Mismatches 150; Indels 20; Gaps 4;
QY 1908 TTGCTTTAGGCAGGAGTTGGGACAGCTGGGCAACATAGGGGATCCCATCTCTACA 1967
Db 127 TAGTATGTCCCGAGGATTCAGACTATCTCTGGGCAACATAGGAGACTCCATCTCT--- 184
QY 1968 CACAAAAAATTTTAAATGAACAGGACATTTGGCATCGGCTATAGTCCCGGACCTC 2027
Db 184 -ACAAAAAATTTAAATTTGAACAGGACATGATGGACATGCTGTGGTCCCGACTACAC 242
QY 2028 AGAGGCACAGCGGGAGAGATCACTTGACCTGGGAGGTTGTGGTGCAGTGCATGCA 2087
Db 243 AGGAGGCTGAGGTGGGAGAGTTCGTTGAGCCCGAGGAGGTGCAGGTGCAGTGCATGA 302
QY 2088 TTGTACCACTGCACCTCGCTGGGCAACAGAGCAAGACTTGTCTCAAAAAATAACAAA 2147
Db 303 TCACACACTGCTCTTCAG-CTTGGCAACAGAGTGGAGCCCTGTCTCTAAAAAGAAAAA 361
```

```
QY 2148 CTAAATTA--AAAAAGAAGACAGAGATAGTGGTGTGGTGGCTCACACTTCAATCC 2205
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 362 AAGAATTACCTAATATATATAAGAAACAAGCCAGGCACAGTGGCTCACCTATATACC 421
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2206 CAGCACTTTGGAAGCCGAGGTGGGCAGATCATCTGAGGCCAGGAGTTCAAGACAGCT 2265
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 422 CAGCTCTTTGGAGGCCAAGCGGGTGGACCACTGAGTCAGGAGTTCAAGACAGCT 481
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2266 GCCTAACATGGTGAATCCTATCTCTACCAAAATACAAAATTAGCCAGCGGTGGT 2325
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 482 GGCCAACATAGCAAAACGCCGTCTCTACTAAAAATACAAAATTAGCCGGTGTATGC 541
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2326 GGGCACCTGTACTGGGAGGTGCCCAACCCAGCTACTCTGGGAGGCTGAGTCAGGAGAACG 2385
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 542 ACGTCCCTGTAT-----CCCCACTACTCAGGAGGCTGAGCGCAGGAGATCA 588
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2386 CTTGAACCTGGGAGCGGAGGTTCGGTCACTGAGATGGTGCCACTCCTCCAGCCTG 2445
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 589 CTTGAACCCAGGAGGTGGAGGTTCAGTGAGCGGAGATTGTGACACTGTACTCCAGCCTG 648
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2446 GCGGAAGAGCGACTCTCTCTCCAAAAAAGAGAGAGGAGGACACAGACACACA 2505
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 649 AGTGACAGCAAGACTCCGCTCTCAAAAAACAAACAAACAAACAAACAAACCCAAA 708
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2506 GA 2507
Db 709 TA 710
```

AA081138 542 bp mRNA EST 21-OCT-1996
zn34h07.s1 Stragene endothelial cell 937223 Homo sapiens CDNA
clone IMAGE:549373 3' similar to contains Alu repetitive element;
mRNA sequence.

AA081138
AA081138
GI623221
AA081138.1 GI:1623221
EST.
human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 542)
Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chiapelli, B.,
Chisoe, S., Dietrich, N., DuBuque, T., Favello, A., Gish, W.,
Hawkins, M., Hultman, M., Kucaba, T., Lacy, M., Le, M., Le, N.,
Mardis, E., Moore, B., Morris, M., Parsons, J., Prange, C., Rifkin, L.,
Rohlfing, T., Schellenberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J.,
Trevisan, E., Underwood, K., Wohldmann, P., Waterston, R., Wilson, R.
and Marra, M.

Generation and analysis of 280,000 human expressed sequence tags
Genome Res. 6 (9), 807-828 (1996)
97044478
On Sep 12, 1996 this sequence version replaced gi:194981.

CONTACT: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -40M13 fwd. from Amersham
High quality sequence stop: 442.
Location/Qualifiers
1. 542
/organism="Homo sapiens"
/db_xref="GDB:3927890"
/db_xref="taxon:9606"
/clone="IMAGE:549373"

FEATURES
source

/clone_lib="Stratagene endothelial cell 937223"
/dev_stage="umbilical vein, 1 passage"
/lab_host="SOLR (kanamycin resistant)"
/note="vector: pBluescript SK-; Site_1: EcoRI; Site_2:
XhoI; Cloned unidirectionally. Primer: Oligo dt.
Umbilical vein endothelial cells, passaged once. Average
insert size: 1.0 kb; Uni-ZAP XR vector; -5' adaptor
sequence: 5' GAATTCGACGAG 3' -3' adaptor sequence: 5'
CTCGAGTTTTTTTTTTT 3"

BASE COUNT 104 a 161 c 121 g 153 t 3 others
ORIGIN

Query Match 5.78; Score 278.6; DB 28; Length 542;
Best Local Similarity 73.08; Pred. No. 4.7e-25;
Matches 408; Conservative 0; Mismatches 126; Indels 25; Gaps 3;

Qy 1885 CTCTGGGAGGCAAGATGGAGGATGCTTGAAGCCAGGAGTTGGACCACTGGGCAA 1944
Db 541 CTCAGGAGCCNAAGTGGAGGATCACTTGGACCCAGGAGTTTGTAGACTGCCCTGGCAA 482

Qy 1945 CATAGGGGATCCCTCTCTACACACAAAAATTTTAAATGAACCAAGGATTTGTGCA 2004
Db 481 CTAAGTGAGACTCCATCTCTAC-----AAAAATTAATAATAGCCAGGTGGAGNG 427

Qy 2005 TGGCGCTATAGTCCAGCCACTCAAGAGGCACAGCGGGAGGATCACTTGAAGCTGGGAG 2064
Db 426 CATCGCTGTGCTCAGCTATTATGAGGCGCTTGAAGCAGGAGATCACTTGAAGCAGGAA 367

Qy 2065 GTTGTGGTTGAGTCACTATGATGTACCACTGCCTCCAGCTGGGCAACAGCAAG 2124
Db 366 GTTGAGGCTGCAATGAGCGGTGATCATGCCACTGCCTCCAGCTGGGTGACAGGTGAG 307

Qy 2125 ACCTGTCTCAAAATAAACAACTAATA-----TAAAAAAGAGAGAGAGATAG 2177
Db 306 ACCCTGTCTTAAAAAAGAAAGAAAGAACGCTGCTTGTAAAGCGTACAGTAGG 247

Qy 2178 TGGGTGTGGTGTACACCTGCAATCCAGCAGCTTTGGAAGCGCGGAGGTGGGAGATCA 2237
Db 246 TCAGCGGTGGTGTGCTCATGCTTATCTAGCAGCTTTGAGAGGCTGAGGTGGGTAGATCA 187

Qy 2238 TCTGAGGCGAGGATTCAGACCAAGCTGGCTAACATGTCGAATCTCTATCTACCAAA 2297
Db 186 CTTGAGGTCAGGGGTTTGAACAGAGCTTGGCCACATCGGCAACCCACCTCTACTANA 127

Qy 2298 AATCAAAAAATTAGCAGCGCTGTGTGGGCACTGTACTTGGGAGGTGCCACCCAGC 2357
Db 126 AATCAAAAAATTAGCGGCGGTGTAGCGCATGCTGTGTA-----ACCCGAGC 80

Qy 2358 TACTGGGAGGCTGAGTCAGAGATCGCTTGAACCTGGGAGGGAGGTTGGCGTACG 2417
Db 79 TACTGGGAGGCTGAGGAGGAGATCGCTTGAACCTAGGAGGGAGGTTGCGAGTGAGC 20

Qy 2418 TGAGATGGTGCACATGCAC 2436
Db 19 CGAGATGGCGCCACTGCAC 1

RESULT 6
LOCUS AL048626/c 540 bp mRNA EST 30-APR-1999
DEFINITION DKFZp564f1822_r1 564 (synonym: hfb2) Homo sapiens cDNA clone
ACCESSION DKFZp564f1822, mRNA sequence.
NID AL048626
VERSION AL048626.1 GI:4727160
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 540)
AUTHORS Wambutt, R., Heubner, D., Mewes, W., Gassenhuber, J. and Wiemann, S.

TITLE EST (Wambutt, et al.)
JOURNAL Unpublished (1999)
COMMENT On Jun 5, 1998 this sequence version replaced gi:3188302.

CONTACT: Wambutt R
MIPS
Am Klopferspitz 18a D-82152 Martinsried, Germany.
Location/Qualifiers
source 1. .540
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="DKFZp564f1822"
/clone_lib="564 (synonym: hfb2)"
/tissue_type="brain"
/dev_stage="fetal"
/lab_host="X1-2blue"
/note="vector: pAMP1; Site_1: NotI; Site_2: SalI"

BASE COUNT 90 a 146 c 125 g 179 t
ORIGIN

Query Match 5.6; Score 274; DB 49; Length 540;
Best Local Similarity 72.88; Pred. No. 1.7e-24;
Matches 396; Conservative 0; Mismatches 140; Indels 8; Gaps 3;

Qy 1613 CAACATGGTAAACCCCATCTCTACCAAAAAATACAAAAATTAGTGGCGTTGGTGGCTGG 1672
Db 539 CAACATGGCAAAACCCCTGTATCTACTGAAATACAAAAATTAGTGGCGATCGGTGCA 480

Qy 1673 CGCTGTAAATCCAGCTACTCAGGAGACTCAGGAGAGATCGCTTGAACCTGGGAGGC 1732
Db 479 TGCCTGTAAATCCAGCTACTCAGGAGGCTCAGGAGAGATCGCTTGAACCTCAGGAGGC 420

Qy 1733 AGAGTTTGCAGTGAAGCGAGATTGGCCACTGGCACTCCAGCTCCGCTGACAGAGAGCCTG 1792
Db 419 AGGGTTTGCAGTGAAGCGAGATCGCACCTGGCACTCCAGCTCCGCTGACAGAGAGC 360

Qy 1793 TCTCTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1849
Db 359 TCCGTTTCAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 300

Qy 1850 GCCAGGTGAGTGGCTCAGCCCTATATCTCAGCACCTCTGGAGGCCAAGATGAGGAGTT 1909
Db 299 GCTGGGTGGGTCAGCCCTCTCTGTAATCCAGCACCTTTGGAGGCTGAGGAGCAG-GCAGA 241

Qy 1910 GCTTGAAGCCAGGAGTTTGGGACAGCCTCGGCAACATAGGGGATCCCATCTCTACACA 1969
Db 240 TCACGAGTTCAGGAGATTGAGACCATCTGGCTAACACGCTGAAACCGAGTCTCTACTAA 181

Qy 1970 CAAAAAATTTTAAATGAACCCAGGCAATTTGGCATCGGCTATAGTCCAGCAGCTCAA 2029
Db 180 AAATACAAA-----AAATGAGCCGGGTATGGTGGCGGCACTGTAGTCCAGCTACTCGG 125

Qy 2030 GAGGCACAGCGGGAGGATCACTTGAAGCTTGGAGGTTGGTTGTCAGTGAAGCTATGATT 2089
Db 124 GAGGCTAGGCGAGGAGATGCGGTGAACCTTGGGAGCGGAGCTTGCAGAGAGCTGAGATC 65

Qy 2090 GTACCACTGCACCTCCAGCCCTGGGCAACAGAGACCTTTGTCTCAAAAAATAAACAAC 2149
Db 64 ACACCACTGCACCTCCAGCCCGGACAGAGAGAGACTCCGCTCTCAAAAAATAAATAA 5

Qy 2150 AAAA 2153
Db 4 AAAA 1

RESULT 7
LOCUS AA680243/c 529 bp mRNA EST 02-DEC-1997
DEFINITION ac86a01.s1 Stratagene ovary (#937217) Homo sapiens cDNA clone
IMAGE:869448 3', similar to contains Alu repetitive element; contains
element L1 repetitive element ; mRNA sequence.
ACCESSION AA680243

Martin, J., Moore, B., Schellenberg, K., Steptoe, M., Tan, F.,
Theising, B., White, Y., Wylie, T., Waterston, R. and Wilson, R.
WashU-NCI human EST Project
Unpublished (1997)
On May 9, 1995 this sequence version replaced gl:802598.

TITLE
JOURNAL
COMMENT

Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu

This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Putative full length read

The vector to vector length is 466
Insert Length: 503 Std Error: 0.00
Seq primer: primer name ambiguous
High quality sequence stop: 431.

FEATURES
source

1. 457
/organism="Homo sapiens"
/db_xref="GB:1261371"
/db_xref="taxon:9606"
/clone="IMAGE:327169"
/clone_lib="Soares_fetal_heart_NbHH19W"
/sex="unknown"
/dev_stage="19 weeks"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: heart; Vector: p7T3D (Pharmacia) with a
modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'
TGTTACCAATCTGAAGTGGGAGCGCGCATCTTTTCTTTTCTTTT 3']
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified p7T3 vector
(Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by
M.Fatima Bonaldo. This library was constructed from the
same fetus as the fetal lung library, Soares fetal lung
NbHL19W."

BASE COUNT 149 a 98 c 123 g 87 t

ORIGIN

Query Match 5.3%; Score 257.2; DB 31; Length 457;
Best Local Similarity 75.8%; Pred. No. 1.7e-22;
Matches 350; Conservative 0; Mismatches 98; Indels 14; Gaps 2;
QY 2022 CCACCTCAAGAGGCACAGCGGGAGGATCACTTGAGCGCTGGGAGGTTGTGGTTCAGTGAG 2081
Db 1 CTACTCAGGAGGCTGAGCGGAGGAGATCACTTCAACCTGGGAGCGGAGGTTGCTAGT 60
QY 2082 CTATGATTGTTACCACTGCAGCTCAGCGTGGGCAACAGCAGACGCTTGTCTCAAAAATA 2141
Db 61 TCGAGATCATGACTGCAGCTCAGCGTGGTAAACAGCAGACGCTCCGTCTCAAAAAGA 120
QY 2142 AACAACTAAATTAATAAAGAGACGAGATAGTGGTGTGGTGCACACCTGCA 2201
Db 121 AA-AAAAAAGAGGATGTTTAAATCTGGCGGCGATGATGGCTCACACTGTA 179
QY 2202 ATCCAGCACTTTGGAAGCCGAGGTGGGAGATCATCTGAGCCAGGAGTTCAGACCA 2261
Db 180 ATCCAGCACTTTGGGAGGCTGAGCGGGCTGATCACTTGAAGTTCAGGAGTTCAGACCA 239
QY 2262 GCTGCTAACATGTTCAATCTTCTATCTACCAAAAATACAAAATAGCCAGGCGTGG 2321
Db 240 GCTGCTAACATGTTGAAATCCGTTTCTTACTAAAAATACAAAATAGCCAGTGTGG 299
QY 2322 TGGTGGGCACCTTACTTGGGGAGGTGCCACCCAGCTACTTGGGGAGGCTGAGTCAGGAGA 2381
Db 300 TGGCAGACACCTGTAT-----CCAGCTACTTGGAAAGGCTGAGGCAGGAGA 346

QY 2382 ATCGCTTGAACCTGGGAGCGAGGTTGGGTACGTGAGATGGTCCACCTGCACCTCCAG 2441
Db 347 ATCACTTGNACCAAGAGGTGGAGGTTCAGTACGTGAGATGGTCCACCTGCACCTCCAG 406
QY 2442 CCGTGGCGGAAGAGCGACTCTGTCTCCAAAAAAGAGAAGA 2483
Db 407 CCTGGCAACAGAGTGAGACTCCATTTCAAAAAAAGAAAAA 448
RESULT 11
LOCUS AI038990 496 bp mRNA EST 28-AUG-1998
DEFINITION ox24f02.x1 Soares_total_fetus_Nb2HF8.9w Homo sapiens cDNA clone
IMAGE:1657275 3' similar to contains Alu repetitive element;; mRNA
sequence.
ACCESSION AI038990
NID G3278184
VERSION AI038990.1 GI:3278184
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 496)
AUTHORS NC-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On Jan 14, 1998 this sequence version replaced gl:1796960.

Contact: Robert Strausberg, Ph.D.

Tel: (301) 496-1550

Email: Robert.Strausberg@nih.gov

This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.

Insert Length: 915 Std Error: 0.00

Seq primer: -40m13 fwd. ET from Amersham

High quality sequence stop: 473.

FEATURES
source

1. 496
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1657275"
/clone_lib="Soares_total_fetus_Nb2HF8.9w"
/dev_stage="8-9 weeks"
/lab_host="DH10B"
/note="Vector: p7T3D-pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was prepared from mRNA obtained from pooled 8-9 week
(total) fetus material with a Not I - oligo(dT) primer [5'
TGTTACCAATCTGAAGTGGGAGCGCGCTTAATTTTCTTTTCTTTT 3']
Double-stranded cDNA was ligated to Eco RI adapters
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified p7T3 vector. Library
went through one round of normalization, and was
constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 86 a 139 c 129 g 142 t

Query Match 5.2%; Score 255.2; DB 41; Length 496;
Best Local Similarity 77.7%; Pred. No. 2.9e-22;
Matches 390; Conservative 0; Mismatches 93; Indels 19; Gaps 6;

QY 1974 AAAATTTTATGACAGGCGATGTGCATGCGCTATATGCCACCCACTCAAGAGG 2033
Db 486 AAAATACAAAATAGCAGGCGATGTGGCACACCGCTGTATCCCGCTAGGAGG 427
QY 2034 CACAGCGGGAGGATCACTTGAGCCTGGGAGGTTGTGGTTCAGTGAGCTATGTTGAC 2093
Db 426 CTGAGGCGAGGAGATCGCTTGAACCCGGGAGGCGGAGGTTGCCGTGAGCTGAGATCGCGC 367
QY 2094 CACTGCATCCAGCCTGGGCAAC-AGAGCAGACCTTGTCTCAAAAATAAACAAATAA 2152

```

|||||
Db 366 CACTGCACTCAGCTGGGCAACAGCGGAACTCTGTCTCAAAAAAAAAAAAAAAAAA 307
QY 2153 ATTA-AAAAAAGACGAGAGATAGTGGGTGGTGGCTCACACCTGCAATCCCGACAC 2211
Db 306 AGTATATATTTAAAGACATTTGGCGGTGGTGGCTCACGCTGTATCCCGACAC 247
QY 2212 TTTCGAAGCGGAGGTGGGAGATCTCTGAGGCGGAGGTTCAAGACCGCTGGCTAA 2271
Db 246 TTTCGGAGCGGAGGTGGGAGATCA--CGAGGTTCAGGAGATGAGACCATCTGGCCAA 189
QY 2272 CATGTGAATCTCTCTACCAAAAAATACAAAAATAGCAGCGCTGGTGGGCGAC 2331
Db 188 CACGGTAACCCCGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 130
QY 2332 CTGTACTGGGAGTGGCCACCCAGCTACTGGGAGGCTGAGTCAGGAGATCGCTTGAA 2391
Db 129 CTGTAAT-----CCAGCTACTGGAGAGGCTGAGGCGAGGAGATCGCTTGAA 83
QY 2392 CCTGGGAGCGGAGTGGCGTCACTGAGTGGTGGCTGAGTGGCTGAGTGGGCGA- 2451
Db 82 CCCGGAGCGGAGGTTGGCGTGGCTGAGTGGCTGAGTGGCTGAGTGGCTGAGTGGGCG 23
QY 2451 AAGAGCGACTCTGTCTCAAAA 2472
Db 22 AAGAGCGNACTCTGTCTCAAA 1

```

```

RESULT 12
AI679002/c
LOCUS
DEFINITION
tuf6901.x1 NCI_CGAP_Gas4 Homo sapiens cDNA clone IMAGE:2255472 3'
similar to contains Alu repetitive element; contains element TARI
repetitive element ;, mRNA sequence.
ACCESSION
AI679002
NID
94889184
VERSION
AI679002.1 GI:4889184
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 487)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
AUTHORS
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL
Unpublished (1997)
COMMENT
On Mar 10, 1998 this sequence version replaced gi:2948847.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Seq primer: -40UP from Gibco
High quality sequence stop: 422.
Location/Qualifiers
1. 487
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="788H12; 14q24.3"
/clone="IMAGE:2255472"
/tissue_lib="NCI_CGAP_Gas4"
/tissue_type="poorly differentiated adenocarcinoma with
signed ring cell features"
/lab_host="DH10B"

```

```

FEATURES
source
1. 487
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="788H12; 14q24.3"
/clone="IMAGE:2255472"
/tissue_lib="NCI_CGAP_Gas4"
/tissue_type="poorly differentiated adenocarcinoma with
signed ring cell features"
/lab_host="DH10B"

```

```

/note="Organ: stomach; Vector: pCMV-SPORT6; Site_1: SalI;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dr.
Average insert size 1.69 kb. Life technologies catalog #:
11549-011"
BASE COUNT 104 a 139 c 110 g 134 t
ORIGIN
Query Match 5.2%; Score 255.2; DB 50; Length 487;
Best Local Similarity 73.0%; Pred. No. 2.9e-22;
Matches 375; Conservative 0; Mismatches 98; Indels 41; Gaps 2;
QY 1954 ATCCCATCTCTACACAAAAAATTTTAAATGAACCCAGGCATTGTGGCATGGCCTAT 2013
Db 476 ATGAGACCCCTTCTCTACAGAAAATTTAAATAATCAGCCAGCATGGTGATATGTGCCTGT 417
QY 2014 AGTCCAGCCAGCTCAAGAGGCGGAGGAGGATCACTTGAGCGCTGGAGGTTGTGGTT 2073
Db 416 AGTCCAGTTACTCAGAGGCTGAGGCGGAGGATGCTTGAGCCAGGAGCTCAAGGCT 357
QY 2074 GCAGTGAGCTATGATGTACCTCCAGCTGAGCTGGGCAAGAGAGCAAGCTTGTCT 2133
Db 356 ACAGTGAGCTATGATGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 297
QY 2134 CAAAAATAAACAACATAAAATTTAAAAAAGAACAGAGATAGTGGTGTGGCTCA 2193
Db 296 TTAATAAATAAAGCA-----GGTGTGGCTCA 265
QY 2194 CACCTGCAATCCAGCAGCTTTTGAAGCCGAGGTGGGAGCATCATCTGAGGCCAGGAGTT 2253
Db 264 TGCTGTATCCAGCATTTTGGAGGCTGAGGCAAGTGATCACCTGAGGTGAGGATT 205
QY 2254 CAAGACAGCCTGGCTAACATGCTGAAATCTATCTACCAAAAAATACAAAAATTAGCC 2313
Db 204 GAAGACCAGCCTGACCAACATGCTGAAACCGCATCTCTACTAAAAATACAAAAATTAGCT 145
QY 2314 AGCGTGGTGGTGGGACCTGTACTGGGAGGTGCCACCCAGCTACTGGGGAGGCTGAG 2373
Db 144 GGGCATGGTGGAGGATCTGTAAT-----CCAGCTACTTGGGAGGCTGAG 98
QY 2374 TCAGGAGAAATCGCTTGAACCTGGAGGCGGAGGTGGGTGAGTGTGGTGGCTGAGT 2433
Db 97 GCAGGAGACTGTGTAACCCCGGAGGCGGAGGTGAGGTGAGCCGAGATCAAGCCATTG 38
QY 2434 CACTCCAGCTGGGCGGAGAGAGCGACTCTCTCTC 2467
Db 37 CACTCCAGCTGGGTGACAGAGTGAGACTCCATC 4
RESULT 13
AI653525/c
LOCUS
DEFINITION
AI653525 490 bp mRNA EST 04-MAY-1999
tq95e03.x1 NCI_CGAP_Ov23 Homo sapiens cDNA clone IMAGE:2216572 3'
similar to contains Alu repetitive element;; mRNA sequence.
ACCESSION
AI653525
NID
94737504
VERSION
AI653525.1 GI:4737504
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 490)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
AUTHORS
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL
Unpublished (1997)
COMMENT
On May 18, 1998 this sequence version replaced gi:3137414.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.

```

Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

Seq primer: -40UP from Gibco
High quality sequence stop: 411.

FEATURES

Location/Qualifiers
1..490
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2216572"
/clone_lib="NCI_CGAP_Ov23"
/tissue_type="tumor, 5 pooled (see description)"
/lab_host="DH10B"
/note="Organ: ovary; Vector: pCMV-SPORT6; Site_1: SalI;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.35 kb. Tumor types include: mixed
Mullerian tumor, papillary serous, clear cell, spindle
Cell. All are primary tumors, metastasis positive. Life
Technologies catalog #: 11534-013"

BASE COUNT

98 a 132 c 105 g 155 t

Query Match

Best Local Similarity 5.2%; Score 253; DB 49; Length 490;

Mismatches 358; Conservative 0; Mismatches 130; Indels 7; Gaps 2;

QY 1984 AATGAACAGGATGTGGCATGCTATAGTCCAGCCACTCAAGAGGCACAGCGGG 2043
DB 490 AATTAGCCAGGATGCTGTCATGCACCTGAATCCAGGTGCTGTGGAGCTGTAGGTGG 431
QY 2044 AGGATCACTTGAAGCTGGGAGGTGTGGTTCAGTGAGCTATGTTACCACTGCATC 2103
DB 430 AGAATCGATTGAGCCAGGAGGTCTAGGCTACAGTGAGCCATGATCACCACCTGCATC 371
QY 2104 CAGCCTGGCAACAGACAGACCTGTCTCAAAATAAACAATAAATAAATAAAG 2163
DB 370 TAGCCTGAGTGGCAGAGTGAAGCCCTGTCTCAAAACATAAATAAATAAATAAAT 311
QY 2164 AAGACGAGATAGTGGGTGGTGGCTCACACCTGCAATCCAGCACTTTGGAAGGCG 2223
DB 310 A-----AATTATAGGGTGCAGTGGCTCATGCTGTATCTAGCATTGGAAGGCG 256
QY 2224 AGGTGGCAGATCATCTGAGGCCAGAGTTCAAGACCAGCCTGGCTAACATGTTGAATC 2283
DB 255 AGGCGGCGAGATCACTTGAGTTCAGAGTTTGAGACCAGCCTGGCAACATGGCGAAGCC 196
QY 2284 CTATCTCTACCAAAATACAAAATATAGCCAGCGGTGGTGGTGG--GGCACTGTACTGGG 2341
DB 195 CTGTCTCTACTAAAATACAAAATATAAATAAATAAATAAAGAGCTTGGTGTGGCAGG 136
QY 2342 GAGGTGCCACCCAGCTACTGGGAGGCTGAGTTCAGGAGAAATCGCTTGAACCTGGGAGGC 2401
DB 135 TGCCTGTAATCTCAGCTACTCAGGAGCTGAGGAGGAGAACTTGAACCCAGGAGGC 76
QY 2402 GGAGTTCGGGTGAGCTGAGATGGTGGCCACTGCATCTCCAGCTGGCGCAAGAGCACTTC 2461
DB 75 AGAGATTGCAATGAGCTGAGATCAGGACACTGCACCTCCAGCCTGGGTGACAGCAAGAC 16
QY 2462 TGCTCTCCAAAAA 2476
DB 15 TGCTCAAAAAA 1

RESULT 14

A1054414/c

LOCUS

A1054414 523 bp mRNA EST 27-JUL-1998

DEFINITION

q176h03.x1 NCI_CGAP_Ov26 Homo sapiens cDNA clone IMAGE:1862453 3'

similar to contains Alu repetitive element; contains element THR
repetitive element ;, mRNA sequence.

ACCESSION

A1054414

NID

93222201

VERSION

A1054414.1 GI:3322201

KEYWORDS

EST.

SOURCE

human.

ORGANISM

Homo sapiens

REFERENCE

1 (bases 1 to 523)

AUTHORS

NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

TITLE

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

JOURNAL

Tumor Gene Index

COMMENT

Unpublished (1997)

On Jan 9, 1998 this sequence version replaced gi:935961.

Contact: Robert Strausberg, Ph.D.

Tel: (301) 496-1550

Email: Robert.Strausberg@nih.gov

Tissue Procurement: Monica Brown, M.D., Elise Kohn, M.D., Michael

R. Emmert-Buck, M.D., Ph.D.

cDNA Library Preparation: David B. Krizman, Ph.D.

DNA Sequencing by: Washington University Genome Sequencing Center

Clone distribution: NCI-CGAP clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

www-bio.llnl.gov/bbrp/image/image.html

Insert Length: 826 Std Error: 0.00

Seq primer: -40ml3 fwd. ET from Amersham

High quality sequence stop: 445.

FEATURES

source

1..523

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:1862453"

/clone_lib="NCI_CGAP_Ov26"

/sex="female"

/tissue_type="papillary serous carcinoma"

/dev_stage="adult"

/lab_host="DH10B"

/note="Organ: ovary; Vector: pAMP1; mRNA made from

papillary serous ovarian carcinoma, cDNA made by oligo-dt

priming. Directionally cloned. Size-selected on agarose

gel, average insert size 500 bp. Primary library,

non-amplified."

95 a 145 c 122 g 161 t

BASE COUNT

ORIGIN

Query Match 5.2%; Score 252.8; DB 41; Length 523;

Best Local Similarity 73.9%; Pred. No. 5.5e-22;

Mismatches 390; Conservative 0; Mismatches 102; Indels 36; Gaps 4;

QY 1586 AGCCAGGAGTTCACACAGCAGCTGAGCAACATGTAACCCCATCTCTACCAAAATA 1645

DB 523 AGGTGAGGAGTTCAGACACAGCCTGGCCACACACCGGTGAACACAGCTCTCTACT 465

QY 1646 CAAAATTTAGCTGGGCTGGTGGCTGGCCCTGTAATCCAGCTACTCAGGAGACTGAGG 1705

DB 464 CAAAATTTAGCTGGGCTGGTGGCTGGCCCTGTAATCCAGCTACTCAGGAGGCTGAGG 405

QY 1706 CAGAGGAGTCCCTGAACCTGGGAGGAGAGTTCAGTGGAGCCAGATTCGCGCACTGG 1765

DB 404 CAGAGAGATCACTTGAACCCGGAGGTTGAGGTTGAGTGGAGCAAGATCATCGCCACTGT 345

QY 1766 ACTCCAGCCTGCGTGCACAGAGAGGCTGTCTCTAAATTAATTAATTAATTAATTC 1825

DB 344 ACTCCAGCCTGCGGACAGAGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 314

QY 1826 ATTTTAAAGACGAAAGTACCGCCAGGTCAGTGGCTCAGCGCTATATATCTCAGCAC 1885

DB 1826 ATTTTAAAGACGAAAGTACCGCCAGGTCAGTGGCTCAGCGCTATATATCTCAGCAC 1885

This Page Blank (uspto)

OM nucleic - nucleic search, using sw model
Run on: October 29, 1999, 00:27:33 ; Search time 164.19 Seconds
(without alignments)
7458.982 Million cell updates/sec
GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.

Title: US-09-371-333-1
Perfect score: 4895
Sequence: 1 CTCGCCGCGCTGGCTGGCA.....AAAAAAAAAAAAAAAAAAAA 4895

Scoring table: IDENTITY_NUC

Searched: 311585 seqs, 125096042 residues

Database : N_Geneseq_36.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	642	13.1	2588	1 V40372	Mouse G-protein co
2	618.2	12.6	1080	1 V40373	Mouse G-protein co
3	458.6	9.4	16891	1 X37084	MEFV gene sequence
4	448	9.2	1470	1 Q47355	Myotonic dystrophy
5	420.2	8.6	14796	1 V27941	Survivin gene. Mod
6	402	8.2	32367	1 V35620	Human SHOX (short
7	381.2	7.8	53577	1 T18551	Human polycystic k
8	381.2	7.8	53577	1 T94108	Human PKD1 locus b
9	381.2	7.8	53577	1 T94101	Human PKD1 locus b
10	366.6	7.5	6901	1 T78853	Human lecithin-cho
11	349	7.1	32367	1 V35620	SHOX gene prelinin
12	345.8	7.1	15577	1 V35616	SHOX gene prelinin
13	339.4	6.9	5543	1 T75284	Nucleotide sequenc
14	336.4	6.9	8342	1 Q75209	ALL-1 (acute lymph
15	331.2	6.8	7146	1 V38933	Nucleic acid seque
16	324.6	6.6	8391	1 T16333	MLL gene 8.3 kb fr
17	323	6.6	8392	1 Q53478	MLL gene 8.3 kb Ba
18	321.6	6.6	13747	1 Q47342	Myotonic dystrophy
19	314.4	6.4	22481	1 T11658	Human kidney amino
20	310.4	6.3	45546	1 X23320	GDP-Fuc:beta-D-gal
21	306.6	6.3	8174	1 Q13332	DNA encoding a gly
22	306.6	6.3	8174	1 Q56908	Human alpha(1,2)-f
23	306.6	6.3	8174	1 T61677	Human alpha(1,2)-f
24	306.6	6.3	3373	1 T76768	Genomic clone enco
25	303.8	6.2	10475	1 N80616	Sequence of human
26	303.8	6.2	10475	1 N80643	SHOX gene exon Vb
27	303.4	6.2	625	1 V35615	Human transcriptio
28	303.4	6.2	1354	1 V35619	FMF associated pro
29	303	6.2	3512	1 X37085	Human Cdn-2 DNA. N
30	302.4	6.2	6511	1 Q95493	Human flavin-conta
31	301.6	6.2	25464	1 V57274	Hereditary haemoch
32	301.6	6.2	23736	1 V57903	DNA encoding macro
33	296.6	6.1	2923	1 T76529	Coding sequence fo
34	296.6	6.1	2923	1 T99233	Macrophage derived
35	296.6	6.1	2923	1 X32817	BRCA1, human breas
36	295.6	6.0	24026	1 T18325	Mutated BRCA1 geno
37	295.6	6.0	24025	1 T17455	Mutated BRCA1 geno
38	295.6	6.0	24026	1 T17512	Mutated BRCA1 geno
39	295.6	6.0	24026	1 T17513	Mutated BRCA1 geno
40	295.6	6.0	24026	1 T17514	Mutated BRCA1 geno
41	295.6	6.0	24025	1 T17515	Mutated BRCA1 geno
42	295.6	6.0	24026	1 T17516	Mutated BRCA1 geno
43	295.6	6.0	24026	1 T17517	Mutated BRCA1 geno

ALIGNMENTS

RESULT 1

V40372

ID V40372 standard; cDNA; 2588 BP.

AC V40372;

DT 16-OCT-1998 (first entry)

DE Mouse G-protein coupled receptor 69A08 #1 encoding cDNA.

KW Mouse; CKDLR201.1; chemokine; G-protein coupled receptor; GPCR;

KW 7 transmembrane receptor; inflammation; asthma; antiviral;

KW abnormal cell proliferation; regeneration; degeneration; atrophy; ss.

OS Mus sp.

FH Key

FT Location/Qualifiers

FT 1..1083

FT /tag= a

FT misc_feature

FT 158

FT /tag= b

FT /note= "Residues 158, 159 and 276 are probably absent,

FT changing the reading frame between those

FT positions; sequence are provided in V40373 and

FT W69598"

FT PN WO9831810-A2.

PD 23-JUL-1998.

PR 20-JAN-1998; U00218.

PR 21-JAN-1997; US-786624.

PI (SCHE) SCHERING CORP.

PI Capone M, Gorman DM, Hedrick JA, Huffine CF, Rossi DL,

PI Vicari A, Zlotnik A;

DR WPI; 98-414108/35.

DR P-PSDB; W69597.

PT Pure or recombinant chemokine CKDLR201.1 - useful, e.g. for treating

PT inflammation and as antiviral agents

PS Claim 7; Page 61-64; 77pp; English.

PS The present sequence encodes mouse G-protein coupled receptor 69A08 #1

CC protein which comprises a plurality of epitopes. Host cells containing

CC vectors comprising a nucleotide sequence encoding the protein are used to

CC produce recombinant protein. Treatment of a cell, particularly neuron,

CC macrophage or lymphocyte, carrying a G-protein coupled receptor (GPCR)

CC or a receptor responsive to CKDLR201.1 chemokine protein, with an

CC antagonist is used to control physiological development, e.g.

CC alteration of calcium ion influx, a chemotactic response, morphology,

CC sequences the CKDLR201.1 protein, or GPCRs, are useful as primers or

CC probes, e.g. for detecting and isolating related sequences and for

CC expressing antigenic peptides. Antibodies (Ab) directed against the

CC CKDLR201.1 protein and GPCRs are used to detect or purify the proteins;

CC diagnostically (e.g. for developmental abnormalities); in screening for

CC potential drugs; to inhibit chemokine/receptor activation; (when coupled

CC to a toxin or radioisotope) for killing specific cells, and to raise

CC anti-idiotypic antibodies. CKDLR201.1 protein and GPCRs and compounds

CC which bind them can be used to treat inflammation, e.g. asthma; as

CC antiviral agents, and to treat abnormal cell proliferation, regeneration,

CC degeneration and atrophy. Therapeutic agents are administered orally, by

CC injection and rectally.

CC Sequence 2588 BP; 462 A; 782 C; 743 G; 601 T;

Query Match 13.1%; Score 642; DB 1; Length 2588;

Best Local Similarity 74.5%; Pred. No. 1.4e-75;

Matches 876; Conservative 0; Mismatches 290; Indels 10; Gaps 5;

QY 290 AGCGCTCTAATCTGCTGCCCGCGGCTACCGAGGCAAGTCTGTGCAATGACAGT 349

Db 1198 AAGTCTCTCAGACAAAGCCCTAATCCACGAGGCTACCGGCAAAATCTGTGCAACGACAGT 1257

QY 350 GACACCTGTGAGCTCCCGGACAGCTCAGCGCACTGTTCTGGCTGGTGGCCACGAG 409


```
Db 280 G CAGTGGCTGATCTGCTGTTGGCCCTGGTGTGCCACACAGACTGGCTTACCACCTTCGCT 339
QY 590 G G C C A G C G C T G G C C C T T C G G G A G G C C C T G C G C C T G C C A G C G C C G C C A C T A T A T G T 649
Db 340 G G C C A G C G C T G G C C A T T G T G A G G C T G C C T G C G G G T G C C A C A G C T G C C C T A T A T G C 399
QY 650 C A C A T G A T A G C T C A G T G C T G C T G G C C G C G C T A G C C T G A T C G C T A C T A C T G G C C C T G 709
Db 400 C A C A T G A T A G T T C A G T G T G C T G C T G C T G C A G T C A G T T G A C A G A T A C T G G C C C T G 459
QY 710 G T G C A C C G C T G C G G G C C G C C C T G C G T G G C G G C C T G G C C C T G A C T G A C T G C A T G 769
Db 460 G T G C A T C C T T G G G G C C C T G A C T G C T G G T C A A G C C C T A C T A C T G A C T G T G T T G 519
QY 770 G C T G T T G G C T A T G G C G C C C T G C A C T G C C C C T G A C A C T G C A G C G C G A G A C C T T C 829
Db 520 G T G C C T G G C T C T G C A G C C A C C C T T G C C T C T A C T A C T G C A T C G C A G A A C T T C 579
QY 830 C G G C T G G C G C C C G A T C G C G T G C T G C C A T G A C C G C T G C C C C T G G A G C A G C A G G C C 889
Db 580 C - G A T T A C T G C C C G A T C G A T C G T G T C A T G A T C G C T G C C C C T G G C T G A G C A G A A C 638
QY 890 T C C A C T G C A A C C G G C T T C A C T G C T G C G C T G T G G C T G T T T C C - T G C C C C T G C T 948
Db 639 T C C A C T G A G A A G C C C T T A C T G C T G C T G C T G C T G C C T G C C T T C C T T G C C A C T G C T 698
QY 949 G G C C A T C G C T G C T G A C G G G C C A C C C T G C A C A C C T G G C G C C A G C G C C G G C G C T A 1008
Db 699 G G C C A T G G C C T G T G C T A T G A A C C A C C C T T C G T G C A T T G C G G C C A A T G C C A G C G C T A 758
QY 1009 C G G C C A G C C C T G A G C T G A C C G A G T G T G C T G C C C T G C C G C T G C C C T T C T T C G T G C C 1068
Db 759 C A G C C A T G C A C T C A G A C T G A C A G C C C T G T A C T T C T C G C A G T G C C T T C T T T C A C A C C 818
QY 1069 C A G C A A C C T G C T G C T G C T G C A T T A C T C G A C C C G A G C C C C A G C G C C T G G G G C A C C T 1128
Db 819 T A G C A T G T G C T G C T G T G C A C T A T T C A A C C C A G C C C T G A G C C T G G G G C A A C T 878
QY 1129 C T A T G T G C C T A C T G C C C A G C C T G G C C T G A G C A C C C T C A A C A G C T G C G T G G A T C C C T T 1188
Db 879 C T A T G A G C C T A T G T G C C C A G C C T G G C A C T C A G C A C C C T C A A C A G C T G C G T A G A C C C T T 938
QY 1189 C A T C T A C T A C T A C T G T G C G C C A G T T C A G G A C A A G T G G G G C A G C G C T T C T T C C A A G C 1248
Db 939 C A T C T A C T A C T A T G T G C C C A T A G T T C A G G A G A A G T A C G C G C - - - T A T G T T G T G C G 995
QY 1249 G T C C C G G G G A C A C C T G C C T C C A A G G C C T C T G C G A A G G G G C A G C C G G G C A T G G G 1308
Db 996 C C A G C C G A G C C A G C A G C T C C T C A G G C C T C A G G A G G C T G A A G C C G A G G A C T G C 1055
QY 1309 C A C C A C T C C C T T T G C T C C A G T G A 1333
Db 1056 C A T T T G C T C T A C A C T T C T G T G A 1080
```

```
RESULT 3
ID X37084/c
AC X37084;
DT 05-JUL-1999 (first entry)
DE MEV gene sequence associated with familial mediterranean fever (FMF).
KW MEV; protein pyrin; Familial Mediterranean Fever; FMF; human;
OS FMF-associated mutant; hereditary disease; colchicine; ss.
FH Homo sapiens.
FT Key Location/Qualifiers
FT CDS 1011..1450
FT /tag= a
FT /note= "contains introns"
FT FT 1011..1287
FT /tag= b
FT FT /number= 1
FT FT 1288..2806
```

```
FT /tag= c
FT /number= 1
FT 2807..3439
FT /tag= d
FT /number= 2
FT 3440..7812
FT /tag= e
FT /number= 2
FT 7813..8162
FT /tag= f
FT 8163..8591
FT /tag= g
FT /number= 3
FT 8592..8684
FT /tag= h
FT /number= 4
FT 8685..10346
FT /tag= i
FT /number= 4
FT 10347..10577
FT /tag= j
FT /number= 5
FT 10578..11044
FT /tag= k
FT /number= 5
FT 11045..11067
FT /tag= l
FT /number= 6
FT 11068..13002
FT /tag= m
FT /number= 6
FT 13003..13118
FT /tag= n
FT /number= 7
FT 13119..13304
FT /tag= o
FT /number= 7
FT 13305..13337
FT /tag= p
FT /number= 8
FT 13338..13700
FT /tag= q
FT /number= 8
FT /cons_splice= (5'site= Yes, 3' site=No)
FT 13701..13731
FT /tag= r
FT /number= 9
FT 13732..13896
FT /tag= s
FT /number= 9
FT 13897..14450
FT /tag= t
FT /number= 10
FT WO9909169-A1.
FT 25-FEB-1999.
FT 20-AUG-1998; UI7255.
FT 21-AUG-1997; US-056217.
FT (USSH ) US DEPT HEALTH & HUMAN SERVICES.
FT Aksentijevich I, Blake T, Centola M, Collins FS,
FT Deng Z, Doggett NA, Fischel-goodsian N, Gumucio DL,
FT Kastner DL, Liu PP, Pras M, Richards RL, Riecke DO,
FT Sood R;
FT WPI; 99-243555/20.
FT P-PSDB; Y09001.
FT A new gene, MEV, associated with Mediterranean Familial Fever
FT Claim 6; Fig 1; 49pp; English.
FT The present sequence represents a novel genomic nucleic acid sequence
FT (MEV) encoding the protein pyrin associated with Familial Mediterranean
FT Fever (FMF). Host cells transfected with a vector comprising a nucleic
FT acid encoding pyrin or its FMF-associated mutant, operably linked to a
FT functional promoter, are used for the recombinant expression of the
FT protein. The invention provides a method for diagnosing risk of FMF that
```

CC comprises analyzing a patient sample for an amino acid or nucleic acid
CC sequence of pyrin; and (b) correlating detection of a mutated sequence
CC with risk of developing FMF; Diagnostic kits developed in this invention
CC are used to identify and treat individuals at risk from FMF, a hereditary
CC disease prevalent in persons having a non-Ashkenazi Jewish, Armenian,
CC Arab, or Turkish background. Prior art FMF treatment with colchicine is
CC not effective in patients who are colchicine-resistant, and this
CC invention will cover all individuals. 4316 G: 4356 T:
SQ Sequence 16891 BP: 3905 A: 4316 G: 4356 T:

[illegible]

Db	9238	GCCAAAGCGGGCGGATCA--CGAGTTTCAGGAGTTCGAGAAACAGCCTGACCAACATGTTGA	9181
Qy	2280	AATCCTATCTCTACCAAAAATACAAAATTTAGCCAGGCGTGGTGGGGCACCTGTACTG	2339
Db	9180	AACCCCATCTCTACTAAAATCTACAAAATTTAGCCAGGCGATGGTGGCACTCACCTATAAT-	9121
Qy	2340	GGGAGGTGCCCAACCCAGCTACTGGGAGGCGTGAAGTTCAGGAGAAATCGCTTGAACCTGGGAG	2399
Db	9121	-----CCAGCTACTCAGSAGSCTGAGSCAGSAGAGATCGCTTGAACCCGGGAG	9074
Qy	2400	CGGAGGTTGCGGTGAGCTAGATGGTGCCACTGTCAGCCTCGGCGGAAAGAGCGAC	2459
Db	9073	GCAGAGGTTTCAGTGTAGCCCAAGATCACACCAGTGCACCTACAACTGGGTGACAGAGCGAG	9014
Qy	2460	TCGTCTCCCAAAAAAAGAGA	2480
Db	9013	ACTCCGTACAAAAAAGAAA	8993
RESULT 4			
QD	Q47355	standard; cDNA; 1470 BP.	
AC	Q47355;		
DT	09-FEB-1994	(first entry)	
DE	Myotonic dystrophy gene fragment containing deletion polymorphisms.		
KS	Myotonic dystrophy; disease; inherited; autosomal dominant; ss.		
QW	Homo sapiens.		
EH	Key	Location/Qualifiers	
FT	primer_bind	1..22	/*tag- a
FT	primer_bind	463..480	/*tag- b
FT	primer_bind	1449..1470	/*tag- c
FN	W09316196-A.		
PD	19-AUG-1993.		
PF	18-FEB-1993;	CA0068.	
PR	18-FEB-1992;	US-837405.	
PA	(KORN/) KORNELUK R G.		
PA	(MAHA/) MAHADEVAN M S.		
FI	Korneluk RG, Mahadevan MS;		
DR	WPI; 93-272897/34.		
PT	Nucleotide molecule - comprises Myotonic Dystrophy locus of		
PT	chromosome 19Q13, for diagnosis of disease status or risk		
PS	Disclosure; Figure 7; 58pp; English.		
CC	Myotonic Dystrophy is an inherited disease and is an autosomal		
CC	disorder. It shows a marked variability in expression		
CC	ranging from a severe congenital form which is frequently fatal		
CC	to an asymptomatic condition associated with normal longevity.		
CC	An increase in the severity of the disease in successive generations		
CC	has been noted. The Myotonic Dystrophy gene (DM) encodes a protein		
CC	with putative serine-threonine protein kinase activity in normal		
CC	individuals. The increase in the severity of the disease in		
CC	successive generations is related to an increase in the number of		
CC	trinucleotide repeats in the DM gene. CTG repeats of up to 40		
CC	constitute a normal gene whereas repeats in excess of 40, especially		
CC	in excess of 50, constitute a mutant gene. The detection and		
CC	subsequent analysis of the number of the repeats region can be used		
CC	to provide information as to an individual's likelihood of developing		
CC	Myotonic Dystrophy. The sequence contains deletion polymorphisms.		
SQ	Sequence 1470 BP: 418 A; 325 C; 446 G; 281 T.		

	Query Match	9.28;	Score 448;	DB 1;	Length 1470;
	Best Local Similarity	69.8%;	Pred. No. 1.5e-50;		
	Matches 727;	Conservative	0;	Mismatches 265;	Indels 50; Gaps 7;
Qy	1514	GGAGAGGCCAGGCGTGGTGGCTCAGCGCTGTAATCCAGCACCTTTAAGAGGCCAGGCGG	1573		
Db	153	GGCGTGGCGGGCGTGGTGGCTCAGCGCTGTAATCTCAGCACCTTTGGAGAGGCCAGGCGG	212		
Qy	1574	ATGGATCACTTTGAGCCCAAGGAGTGTCAACACCAGAGCGCTGAGCAACATGTTAAACACCCCATCT	1633		

Db	213	GTGGATCAC--GAGGTCAGAGATCGAGACCAATCTTGCTTAACACGGTGAACACCCCACT	270
Qy	1634	CTACCAAAAATAC-AAAAATTAGCTGGGCTTGGTGGCTGGCGCCTGTAATCCCGAGCTACT	1692
Db	271	CTCTTAAAAATACAAAAAATTAGCCAGGCGTGTGGCGGCGCTGTAGTCCCGAGCTACT	330
Qy	1693	CAGGAGACTGAGGCACAAGGATCGCTTGAACTTGGGAGGCAGAGGTTGCGAGTGAGCCGAG	1752
Db	331	CAGGAGGCTGAGCAGGAGAAATGGCGTGAACCGGGAGCAGAGTTTGCAGTGAGCCGAG	390
Qy	1753	ATTGGCGCACTGGACATCCAGCTCGGTGACAGAGAGCGTGTCTCTAAATTAATTAATTA	1812
Db	391	ATCGTGCCACTGCATCTCCAGCTTGGCGACAGAGCAGACTCTGTCTCAGAAAAA	450
Qy	1813	TTAATTTAATTCAATTTTAAAAAGACGAAAGTGAAGCCAGGTGCAGTGGCTCAGCGCT	1872
Db	451	AAAA-----AAAAAATAATAGCTGGACCGCGCGCGCTGTGGCTCATGCGCT	502
Qy	1873	ATAATCTCAGCACTCTGGGAGGCCAAGATGGAGGATTCCTTGAAGCCAGAGNATTGGGAC	1932
Db	503	GTATCCCGACACTTTGGGAGTCCAAGCGCGGTGGGTGATGAGATCAGGAGTTTGGAGAC	562
Qy	1933	CAGCCTGGGCAACATAGGGGGATCCCATCTCTACACAAAAAATTTTAAATGAACCA	1992
Db	563	TAGGCTGGCCACACAGGTGAACCCCGTCTTAC---TAAAAATACAGAAATTAGCTG	619
Qy	1993	GGCATGTGGCATGCGCCTATAGTCCCAGCCACTCAAGAGCAGCAGCGGAGGATCACT	2052
Db	620	GGTGTGTCTCGGGTCATGTAATTCAGTTACTGGGAACTGAGCGCAGGAGAATTGCT	679
Qy	2053	TGAGCCTGGGAGGTTGTGGTTCGATGAGCTATGATTACCACTGCACATCCAGCCTGGG	2112
Db	680	TGAACCTGGGAGCGAGTTGCATGTAGCCAAAGATCATGCCACTACACTCCAGTCTGGG	739
Qy	2113	CAACAGACGAAGACCTGTCTCAAAAATTAACAAACTAAAAATTAATAAAGAACACGAGA	2172
Db	740	TGACAGAGTAGACTCTGTCTCAAAAAAATAAAAAA-----A	780
Qy	2173	GATAGTGGGTGGTGGCTACACCTGCACTGCCAATCCAGCACTTTGGAAGCCGAGTGGGCA	2232
Db	781	AGGGTTGGGCAAGTGGTTCAGCGCTGTAATCCCAAGAACTTTGGGAGGCTGCGCAGGCA	840
Qy	2233	GATCATCTGAGCCAGGAGTTCAAGACCAGCCTGGCTAACATGCTGAAATCCCTATCTCTA	2292
Db	841	GATCACTGGAAGTCAGAGTTCAAGACCAGCCTGCCAACATGCTGGAACCCCTGTGTCTA	900
Qy	2293	CCAAAAATACAAAAAATTAGCCAGGGCTGGTGGTGGGCACCTGTACTGGGAGGTGCCAC	2352
Db	901	CTAAAAATACAAATTTAGCCAGGCTTGGTGGCTGTATGCTGTATG-----	948
Qy	2353	CCAGCTACTGGGAGGCTGAGTCAGAGAAATGCG-----TTGAACCTGGGAGCGGAGGTT	2408
Db	948	CCAGCTACTCAGGAGGCTGAGCAGGAGAAATCGCTTGATTTGAACCTGGGAGGCGAGTTT	1007
Qy	2409	GCAGTCAGCTAGATGTCCTCACTCCAGCCTGGCGAAGAGCGGACTCTGTCTCC	2468
Db	1008	GCAGTGGGCTGGGGTTGTGCCACTGCATCTTAGCTGGGACAGCAGACTCCATCTAA	1067
Qy	2469	AAAAAAGAGAGAGGAGGACACAGAGACACAGAGAAAGAACGCAATGTGCGGCA	2528
Db	1068	AAAAAAGACAACTGGGCTGGGCACAGTGGCTTATATTGTAAATCCAGCACTTGG	1127
Qy	2529	GAGCAGAGATGGAGTGATGC	2550
Db	1128	GAGGCTGAGGTTGGAGGACTGC	1149

RESULT

RESOL
V27941/C

ID V27941 standard; cDNA; 14796 BP.

AC V27941;

DT 25-SEP-1998 (first entry)

DE Survivin gene.

survivin; apoptosis; cellular apoptosis; transplantation; ss;
motor neuron degenerative disease; HIV infection; immunosuppression;
gastrointestinal perturbations; cardiovascular disorder

WO9822589-A2.

28-MAY-1998

20-NOV-1997; U21880.

20-NOV-1997; US-975080.

20-NOV-1996; US-031435.

(UYYA) UNIV YALE.

Altieri DC;

WPI; 98-312475/27.

P-PSDB; W614359.

Modulating apoptosis by controlling the Survivin gene - useful for treating transplant rejection, degenerative disorders and tumours
Disclosure: Fig 10: 108pp: English.

The survivin gene can be used to control apoptosis through modification of the gene. Survivin peptides can be used to inhibit cellular apoptosis, e.g. for enhancing the viability of organs and tissues prior to their transplantation, for preserving the growth of cells in culture or for treating conditions involving abnormal apoptosis, e.g. degenerative diseases such as motor neuron degenerative diseases, HIV infection, dermatological effects of ageing, disorders and diseases such as immunosuppression, gastrointestinal perturbations, cardiovascular disorders, apoptosis related to reperfusion damage, rejection of tissue transplantation and Alzheimer's disease. Agents which block survivin activity can be used to treat e.g. tumours.

Sequence	14796 BP;	3482 A;	3488 C;	3888 G;	3938 T;
----------	-----------	---------	---------	---------	---------

Query Match 8.6%; Score 420.2; DB 1; Length 14796;

Query Match 6.0%, Score 420.2, DB 1
Best Local Similarity 70.6%: Pred. No. 4.8e-47:

Best Local Similarity	70.0%	Freq. NO: 4.0e-47;
Matches	725;	Conservative
Mismatches	0;	Mismatches
Indels	243;	Indels
Gaps	59;	Gaps

OV 1486 CAGTGTGGGTGGTCCCTCATAGATAAGGAGAGGCCAGGCCCTGGTGGCTCAGGCCTGTA 1545

QY 1480 CAGTGTGGGTGGTGTCTCATAGATAGGAGAGGCTTGGTGGCTTACGCTGTGTA 1542

ov 1546 ATCCCAGCACACTTTACACGCCCAAGCCGATCGATCCTTCAGCCCAAGGAGCTTCAACACCA 1605

QY 1546 ATCCCAGCACTTTAAGAGGCCAAGGCGGATGGATCACATTGAGCCCAAGGAGTTCAACACCA 1605

1663
1664
1665
1666
1667
1668
1669
1670
1671
1672
1673
1674
1675
1676
1677
1678
1679
1680
1681
1682
1683
1684
1685
1686
1687
1688
1689
1690
1691
1692
1693
1694
1695
1696
1697
1698
1699
1700
1701
1702
1703
1704
1705
1706
1707
1708
1709
1710
1711
1712
1713
1714
1715
1716
1717
1718
1719
1720
1721
1722
1723
1724
1725
1726
1727
1728
1729
1730
1731
1732
1733
1734
1735
1736
1737
1738
1739
1740
1741
1742
1743
1744
1745
1746
1747
1748
1749
1750
1751
1752
1753
1754
1755
1756
1757
1758
1759
1760
1761
1762
1763
1764
1765
1766
1767
1768
1769
1770
1771
1772
1773
1774
1775
1776
1777
1778
1779
1780
1781
1782
1783
1784
1785
1786
1787
1788
1789
1790
1791
1792
1793
1794
1795
1796
1797
1798
1799
1800
1801
1802
1803
1804
1805
1806
1807
1808
1809
1810
1811
1812
1813
1814
1815
1816
1817
1818
1819
1820
1821
1822
1823
1824
1825
1826
1827
1828
1829
1830
1831
1832
1833
1834
1835
1836
1837
1838
1839
1840
1841
1842
1843
1844
1845
1846
1847
1848
1849
1850
1851
1852
1853
1854
1855
1856
1857
1858
1859
1860
1861
1862
1863
1864
1865
1866
1867
1868
1869
1870
1871
1872
1873
1874
1875
1876
1877
1878
1879
1880
1881
1882
1883
1884
1885
1886
1887
1888
1889
1890
1891
1892
1893
1894
1895
1896
1897
1898
1899
1900
1901
1902
1903
1904
1905
1906
1907
1908
1909
1910
1911
1912
1913
1914
1915
1916
1917
1918
1919
1920
1921
1922
1923
1924
1925
1926
1927
1928
1929
1930
1931
1932
1933
1934
1935
1936
1937
1938
1939
1940
1941
1942
1943
1944
1945
1946
1947
1948
1949
1950
1951
1952
1953
1954
1955
1956
1957
1958
1959
1960
1961
1962
1963
1964
1965
1966
1967
1968
1969
1970
1971
1972
1973
1974
1975
1976
1977
1978
1979
1980
1981
1982
1983
1984
1985
1986
1987
1988
1989
1990
1991
1992
1993
1994
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022
2023
2024
2025
2026
2027
2028
2029
2030
2031
2032
2033
2034
2035
2036
2037
2038
2039
2040
2041
2042
2043
2044
2045
2046
2047
2048
2049
2050
2051
2052
2053
2054
2055
2056
2057
2058
2059
2060
2061
2062
2063
2064
2065
2066
2067
2068
2069
2070
2071
2072
2073
2074
2075
2076
2077
2078
2079
2080
2081
2082
2083
2084
2085
2086
2087
2088
2089
2090
2091
2092
2093
2094
2095
2096
2097
2098
2099
2100
2101
2102
2103
2104
2105
2106
2107
2108
2109
2110
2111
2112
2113
2114
2115
2116
2117
2118
2119
2120
2121
2122
2123
2124
2125
2126
2127
2128
2129
2130
2131
2132
2133
2134
2135
2136
2137
2138
2139
2140
2141
2142
2143
2144
2145
2146
2147
2148
2149
2150
2151
2152
2153
2154
2155
2156
2157
2158
2159
2160
2161
2162
2163
2164
2165
2166
2167
2168
2169
2170
2171
2172
2173
2174
2175
2176
2177
2178
2179
2180
2181
2182
2183
2184
2185
2186
2187
2188
2189
2190
2191
2192
2193
2194
2195
2196
2197
2198
2199
2200
2201
2202
2203
2204
2205
2206
2207
2208
2209
2210
2211
2212
2213
2214
2215
2216
2217
2218
2219
2220
2221
2222
2223
2224
2225
2226
2227
2228
2229
2230
2231
2232
2233
2234
2235
2236
2237
2238
2239
2240
2241
2242
2243
2244
2245
2246
2247
2248
2249
2250
2251
2252
2253
2254
2255
2256
2257
2258
2259
2260
2261
2262
2263
2264
2265
2266
2267
2268
2269
2270
2271
2272
2273
2274
2275
2276
2277
2278
2279
2280
2281
2282
2283
2284
2285
2286
2287
2288
2289
2290
2291
2292
2293
2294
2295
2296
2297
2298
2299
2300
2301
2302
2303
2304
2305
2306
2307
2308
2309
2310
2311
2312
2313
2314
2315
2316
2317
2318
2319
2320
2321
2322
2323
2324
2325
2326
2327
2328
2329
2330
2331
2332
2333
2334
2335
2336
2337
2338
2339
2340
2341
2342
2343
2344
23

Qy 1606 GCCTGAGCAACATGGTAAACCCCATCTCTACCAAAATAC--AAAAATTAGCTGGGCTT 1663

6205 6216

Db 6145 GGTGGGGGCGCTGTAGTCCAGCTACTCGGAGGCTGAGCGAGAAATGGCTTGAC 6086
Qy 1724 CTGGGAGCGAGGTTGCACTAGTGCAGATGCGCCACTCGGACTCCAGCTCGGTGACA 1783
Db 6085 CTGGGGGCGAGCTTGCAGTGCAGGAGATCCCGTGCCTGCACTAGACCTGGGCGACA 6026
Qy 1784 GAGACCTGTCTCTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1843
Db 6025 GAGCAAGACTCCAT-----CTCAATAATAATAATAATAATAATAATAATACTCT 5972
Qy 1844 GTGAGCGGCGAGTGCAGTGGCTCAGCCCTATATCTCAGCACTCTGGGAGCG-CAAGATG 1902
Db 5971 G---AGGCGAGGTGGTGGCTCACACCTATATCTCAGCACTTTGGGAGGCTGAGCGG 5915
Qy 1903 GAGGATGCTTGAAGCGGAGGTTGGGACCGCTGGGCAACATGAGGGGATCCATCT 1962
Db 5914 GTGGATCAGCTGAGTGCAGGATTTCAAAACCGAGCTGGCAACATGGTGAACCTCGTCT 5855
Qy 1963 CTACACAAAAAATTTTAAATGAACAGGATGTTGGCATGGCCCTATAGTCCCGAGC 2022
Db 5854 CTAC-----CAAAATACAAAAATACCTGGTGGTGGCACATGCCCTGTAATCTCGGC 5800
Qy 2023 CACTAAGAGGCGAGGCGGAGGATCACTGAGCTGGGAGGTTGTTGTCAGTGGAGC 2082
Db 5799 TACTTGGGAGGCTGAAGCGCAATCACTTGAAGGCGGAGGTTGCACTGATC 5740
Qy 2083 TATGATTGATCCACTGCACTCCAGCTGGGCAACAGAGCAAGCACTTGTCTCAAAAAATA 2142
Db 5739 CGAGATCAGCCACCCCTACTCCAGCTGGTAAACAGATGTAACCTGTCTCAGAACAA 5680
Qy 2143 ACAAACTAAATTAATAAAGAGAGAGATAGTGGTGGTGCATCACACCTGCAA 2202
Db 5679 AAAAA-----AAAAAGCCAGCGTGGTGGTTCACGCGCTGTAA 5644
Qy 2203 TCCAGCACTTTGGAGCGGAGGTGGCGAGATCATCTGAGGCCAGGAGTTCAAGACCAG 2262
Db 5643 TCCAGCACTTTGGAGCGGAGGTGGCGGAATCA- -CGAGGTGAGGAGTTCAAGACCAG 5586
Qy 2263 CTGGCTAACATGTAATCTCTCTACCAAAATACAAAAATACCAAGCGGTGT 2322
Db 5585 CTGGGCAATGTTGTAACCCCTCTCTCTAAATAACAAAAATAGTGGTGGT 5526
Qy 2323 GTGGGCACTGTACTGGGAGGTGCCACCGACTACTGGGAGGCTGAGTCAGAGAA 2382
Db 5525 GGCATTCCTGTAAT-----CCAGTTACTTGGGAGGCTGAGGAGGAGAA 5479
Qy 2383 TCGCTTGAACCTGGAGCGGAGGTGGGTCAGCTGAGATGTCCTGAGTGCACCTCCAGC 2442
Db 5478 TCATTGAACCGAGGCGAGGTTGCAAGTGAAGCAAGATTGCACCTTGCACCTCCAGC 5419
Qy 2443 CTGGCGGA-AAGAGCACTCTCTCTCCAAAAAAGAGAGAGGAGGACACAGAGACA 2501
Db 5418 CTGGCAACAAGAGTGAACCTCTATCTCAGAAAAAATAAATAAATGCTTCTGTGAGGAAA 5359
Qy 2502 CACAGAG 2508
Db 5358 CACAGAG 5352

RESULT 6

V35620/C
ID V35620 standard; DNA; 32367 BP.
AC V35620;
DT 07-SEP-1998 (first entry)
DE Human SHOX (short stature homeobox containing gene) gene sequence.
KW Homeobox domain; human growth gene; growth regulation; growth defect;
KW turner's syndrome; short stature homeobox containing gene; SHOX;
KW SHOX; bone disease; osteoporosis; calcium regulation; short stature;
KW transcription factor A; ss.
OS Homo sapiens.
PN WO9814568-A1.
PD 09-APR-1998.
PF 29-SEP-1997; E05355.

PR 16-JAN-1997; EP-100583.
PR 01-OCT-1996; US-027633.
PA (RAPP/) RAPPOLD-HOEERBRAND G.
PI Rao E, Rappold-hoeerbrand G;
DR WPI; 98-271719/24.
PT New human growth genes - used to develop products for the diagnosis
PT and treatment of human growth defects such as short stature, e.g.
PS Turner's syndrome
PS Claim 19; Pages 51-67; 84pp; English.
CC This is the human SHOX gene sequence containing the PARI region. The gene
CC region corresponding to short stature has been identified as a region of
CC approximately 500 kb in the PARI region of the x and y chromosomes. Three
CC genes in this region have been identified as candidates for the short
CC stature gene. These genes were designated SHOX (also referred to as
CC SHOX93 or HOX93), pT92 and SHOT (SHOX-like homeobox gene on chromosome
CC three). The SHOX gene has two separate splicing sites resulting in two
CC variations SHOXa and SHOXb. The specification provides sequences of SHOX
CC (short stature homeobox-containing) genes SHOX ET92, SHOXa, SHOXb, SHOT
CC and exons of the SHOX genes as shown in V35610 to V35621 and protein
CC sequences of the human growth protein transcription factor SHOXa, SHOXb
CC and SHOT as shown W60573 to W60575. The novel genes are responsible for
CC human growth. Defects in the genes can cause short stature, e.g.
CC turner's syndrome. The products can be used to develop agents for the
CC treatment of short stature or other human growth disorders. The products
CC can also be used for providing a mitogenic effect on cells, e.g. for the
CC treatment of bone diseases such as osteoporosis and diseases involved
CC with disturbance in the bone calcium regulation.
SQ Sequence 32367 BP; 7627 A; 8130 C; 8564 G; 8043 T;

Query Match 8.2%; Score 402; DB 1; Length 32367;
Best Local Similarity 70.1%; Pred. No. 1e-44;
Matches 680; Conservative 0; Mismatches 230; Indels 60; Gaps 8;

Qy 1517 GAGGCCAGCCGTGTGGCTCAGGCTGTAAATCCAGCACTTTAAGAGCCAGGCGGATG 1576
Db 29132 GAGACCACGGGCGAGTGTTCATGTGTAAATCTTAACACTTTGGGAGGTGGAGTGAG 29073
Qy 1577 GATCACTTGAGCCAGGAGTTCACACCCAGCTGAGCAACATGTTAAACCCCTCTCTA 1636
Db 29072 GATCTTTGAGCTCAGGAGCTCAAGACCAGGCTCAGCAGCAACAGATCCCGTCTCTA 29013
Qy 1637 CCAAAAT-ACAAAAATPAGCTGGGCTTGGTGGCGCTGTAAATCCAGCTACTCAG 1695
Db 29012 GTAAAAATCAAAAAATAGCTGGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 28953
Qy 1696 GAGACTGAGGCGAGAGGATCGCTTGAACCTGGGAGGAGAGGTTCCAGTGAGCGGAGAT 1755
Db 28952 GAGGCTGAGGTGGGAGGATTCCTTGAGCCAGGAGTTTGAGGCTGCCAGTGATGATC 28893
Qy 1756 GCGCCACTTGGACTCCAGCTCGCTGACAGAGAGGCTGTCTCTAAATTAATTAATTA 1815
Db 28892 GCACCACTGTCTCCAGCTCGGCTGGGTGACAGAGCAAGACTCTGT-----CTTCTTA 28843
Qy 1816 ATTTAATTCAAATTTAAAGACGAAAGTAGTCAGCGCCAGGTGCTGAGTGGCTCACCGCTATA 1875
Db 28842 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 28783
Qy 1876 ATCTCAGCACTCTGGGAGGCCAA-GATGGAGGATGCTTGAAGCCAGGAGTTGGGACCA 1934
Db 28782 ATCCAGCACTTTGGGAGGCTTAAGCGGGTGGATCACCTGAGGTGAGGAGTTTCAGAGCCA 28723
Qy 1935 GCCTGGGCAACATAGGGGATCCCTCTCTACACACAAAAAATTTTAAATGAACACAGG 1994
Db 28722 GCCTGGCCAACTGGTCAACCCCTCTCTACTGAAATACAA-----ACATTAATCTTGA 28668
Qy 1995 CATGTGCGATCGCGCTATAGTCCAGCACTCAAGAGGACAGCGGGAGGATCACTTG 2054
Db 28667 CATGTGTTGGGCACTCTAATCCAGCTACTCGGGAGGCTGAGGAGGAGAAATCGCTTG 28608
Qy 2055 AGCCTGGGAGGTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 2114
Db 28607 AACTCGGGAGGCGGAGGTTGCGAGTGCAGATCATGCCATTATATCTCCAGCCTGGGCG 28548

QY 2115 ACAGAGCAAGACCTTGTCTCAAAATTAACAACAATAAAATTAATAAAAGAGAGAGAGA 2174
 DB 28547 ACAGAGCAAACTCCATCTATAAAAAAAGAGGCGCA----- 28508
 QY 2175 TAGTGGGTGGTGGCTCACACCTGCAATCCAGCAGCTTTGGAGGCGGAGGTGGGCGAGA 2234
 DB 28508 -----GTGAGTGGCTCACGCTGTAGTCCAGCACTTTGGAGGCTGAGGCGGAGCGGA 28455
 QY 2235 TCATCTGAGGCGAGGAGTCAAGACACCAAGCTGGCTAATACATGGTGAATATCTATCTTACC 2294
 DB 28454 TCA--TGAGTTCAGGAGATCGAGACCTTCTGGCTAATACATGGTGAATATCTTACT 28397
 QY 2295 AAAATATAC-AAAATATAGCAGGCTGGTGGTGGGACCTGTACTGGGAGGTGCCCAAC 2353
 DB 28396 AAAAATACAAAAATAGCGGCTGGTGGGCGGCTGTAGT-----CC 28350
 QY 2354 CAGCTACTGGGAGGCTGAGTCAGGAGAAATCGCTTGAACCTGGGAGCGGAGGTTCGGGT 2413
 DB 28349 CAGCTACTGGGAGGCTGAGGAGGAGAAATGGCTGAACCGGGAAGAGAGCTTGCAGT 28290
 QY 2414 CAGCTGAGTGGTGGCTGCACTCCAGCTCCAGCTGGGCGAAAGAGCGACCTGTCTTCAAAA 2473
 DB 28289 GAGCGGAGATCGTCCACTGCACTCCAGCTGGGTGACAGAGCGAGACTCCGCTCTCAAA 28230
 QY 2474 AAAGAGAGA 2483
 DB 28229 AAAAAA 28220

RESULT 7

ID T18551 standard; DNA; 53577 BP.
 AC T18551;
 DT 06-MAY-1997 (first entry)
 DE Human polycystic kidney disease normal PKD1 gene.
 KW Adult onset polycystic kidney disease; APKD; autosomal dominant;
 KW mutant; transversion; transition; deletion; insertion; ds.
 OS Homo sapiens.
 FH Key
 FT misc_feature
 FT 4379..5272
 FT /tag= a
 FT /note= "specifically claimed region of intronless
 cDNA identified by exon trapping"
 FT old_sequence
 FT replace(50652..50653, cg)
 FT /tag= b
 FT /note= "changes Val codon to Leu codon"
 FT old_sequence
 FT replace(50796..50797, cg)
 FT /tag= c
 FT /note= "replaces Val codon by Leu codon"
 FT old_sequence
 FT insert(51827..51828, cc)
 FT /tag= d
 FT /note= "insertion, results in frameshift"
 FT WO9612033-A1.
 FT 25-APR-1996.
 FT 11-OCT-1995; UI3357.
 FT 12-OCT-1994; US-323443.
 FT 31-JAN-1995; US-381520.
 FT (IGIG-) IG LAB INC.
 FT (UJO) UNIV JOHNS HOPKINS.
 FT Burn TC, Connors TD, Dackowski W, Germino G, Klinger KW;
 FT Landes GM, Qian F;
 FT WPI; 96-222017/22.
 FT Isolated human polycystic kidney disease gene and its mutants
 FT useful for treatment of polycystic kidney disease and screening for
 FT carriers
 PS Claim 1: Fig 1: 65pp; English.
 CC The present sequence is that of the normal human PKD1 gene from
 CC chromosome 16. Mutations in this gene (e.g. transitions,
 CC transversions, deletions and/or insertions) are associated with
 CC adult-onset polycystic kidney disease (APKD). The PKD1 locus is
 CC GC-rich (62.4%). Comparison of this sequence with a previously
 CC reported partial cDNA sequence revealed differences at three

CC locations (see features table). The most significant difference is
 CC the presence of two additional cytosine residues on the plus-strand
 CC at position 4566 of the previously reported sequence. The insertion
 CC results in a frame-shift in the predicted protein coding sequence,
 CC leading to replacement of 92 C-terminal amino acids with a novel
 CC 12 amino acid C-terminus. The PKD1 gene contains 23 Alu repeats.
 CC There is a region consisting of 17 tandem copies of a perfect 27 bp
 CC repeat and two large C-rich regions.
 SQ Sequence 53577 BP; 8495 A; 17681 C; 15785 G; 11616 T;

Query Match 7.8%; Score 381.2; DB 1; Length 53577;
 Best Local Similarity 67.3%; Pred. No. 4.6e-42;
 Matches 719; Conservative 0; Mismatches 228; Indels 121; Gaps 8;

QY 1534 CTCACGCTGTAAATCCAGCAGCTTTAAGAGCGCAAGCGGATGGATCACTTTGAGCCGAGG 1593
 DB 7787 CTCACACCTGTAAATCCAGCAGCTTTGGGAGCGCGAGGAGAGATCACCTAAGGCCAAG 7728
 QY 1594 AGTTCAACACACGCTGTAGCAACATGGTAAACCCCATCTCTTACCAAAATA----CAAA 1649
 DB 7727 AGTTCAAGACACGCTGTAGCAACATGGTAAACCCCATCTCTTACCAAAATA----- 7668
 QY 1650 AATTAGCTGGGCTTGGTGGCTGGCTGTAAATCCAGCTACTCAGGAGACTGAGGCAGA 1709
 DB 7667 AATTAGCGGGCGTGGTGGCGCGGCTCTAATCCAGCTACTCAGGAGGCTGAGGCAGG 7608
 QY 1710 AGGATCGCTTGAACCTGGGAGGAGAGGTTGSCAGTGAGCGGAGATTTGCGCCACTGGACTC 1769
 DB 7607 AGAATCGCTTGAACCGGGAGATGGAGGCTGCAGTGAGCGGAGATCACACATTTGTCTC 7548
 QY 1770 CAGCTGCGTGACAGAGAGGCTGTCTTAATTAATTAATTAATTAATTAATTAATTAATTA 1826
 DB 7547 CAGCTGCGTGACAGAGGAGGCTGTCTTAATTAATTAATTAATTAATTAATTAATTAATTA 7488
 QY 1826 ----- 1826
 DB 7487 CTGGTCCAAGCACAGTGGTGTTCACAAACAAATGTATCACAGCCAGGTAGAAATTTCTCAT 7428
 QY 1826 -----ATTTTAAAGAGCAAGTAAAGTACAGCGC 1851
 DB 7427 TCTTTCTCCAGTCCTACTGCTTTGTTGACAGCCTTAAAGACACACATATATTTTGTG 7368
 QY 1852 CAGGTGCGAGTGGCTCAGCGCTTAAATCTCAGCAGCTCTGGGAGGCCAA--GATGGAGGATG 1910
 DB 7367 TGGGCGGCTTGGCTCAGCAGCTGTAAATCCCAACACTTTTGGGAGGCCAAGGCGGCGATCA 7308
 QY 1911 CTGAAGCCAGGAGTGTGGGACAGCCTGGGCAACATAGGGGATCCATCTCTACACAC 1970
 DB 7307 CTTGAGGTCAGGAGTGTGAGCAGCAGCTGACCAACGTTGGAGAAACCCCGTCTCTCTAAA 7248
 QY 1971 AAAAAATTTTTTAATGAACAGGAGATTTGGCATGCGCTATAGTCCAGCCACTCAAG 2030
 DB 7247 AATACAA-----AATTAGCCAGGAGTGGTGGCAGATGCTGTAAATCCAGCTACTGGAG 7194
 QY 2031 AGGCACAGCGGGAGGAGTACCTTGAAGCTGGGAGGTTGTGGTTCAGTGTAGCTATGATG 2090
 DB 7193 AGGCTGAGCGAGGAGATACCTTGAACCGGAGGCGGAGGTTGCGCGTGTAGTGTAGATCG 7134
 QY 2091 TACCAGTGCAGCTCCAGCTGGGCAAC--AGAGCAAGAGCTTGTCTCAAAAAATAAACT 2149
 DB 7133 CGGCACTGCAGCTCCAGCTGGGCAACAAAGAGCGGAACTCGCTCTCAAAAAAATAAA-- 7076
 QY 2150 AAAATTAATAAAGAGAGAGAGATAGTGGGTGTGGTGTGCTCAGCTGCACTGCAATCCAGC 2209
 DB 7076 AAAGTATATATTTTAAAGAGACATTTGGCGGGTGGCTCAGCGCTGTAAATCCAGC 7017
 QY 2210 ACITTTGAAGCGCGAGGTGGGAGATCATCTGAGGCCAGGAGTTCAGAGCAGCGCTGGCT 2269
 DB 7016 ACTTTGGAGCGCGAGGTGGGAGATCA--CGAGGTTCAGGAGATTCGAGACCATCTCTGCC 6959
 QY 2270 AACATGTGAAATCTTATCTTACCAAAAAATACAAAAATAGCCAGCGCTGGTGGTGGC 2329

[illegible]

Db	7546	CAGCCTGGGTGACAGAGCCAGACTCCGCTCTCAAAACAAAACAAAAGAGACATCAGCTA	74787
Qy	1826	-----	1826
Db	7486	GCTGGTCCAAAGCAGCAGTGGTGTTCACACGAATTCATCACAGCCAGGTAGAAATTCCTTCAT	74277
Qy	1826	-----ATTTTAAAAAGACGAAAAGTGACGGC	1851
Db	7426	TCATTCTCCAGTCTCTACTGTTTGGTTGACCAAGCCTTAAAGACACACATATATTTTGTG	7367
Qy	1852	CAGTGCAGTGGCTCAGGCTATATCTCAGACACTCTGGGAGGCCAA--GATGGAGGATTG	1910
Db	7366	TGGCGCGTTTGGCTCACACTGTAATCCCAACACTTTGGGAGGCCAAGCAGCGGATCA	7307
Qy	1911	CTTGAGCCAGGAGTTTGGGACCAAGCCTGGGCAACATAGGGGATCCCATCTCTACACAC	1970
Db	7306	CCTGAGGTCAGGAGTTTGAGACAGCCTGACCAAGCTGGAGAACCCCGTCTCTCTTAAA	7247
Qy	1971	AAAAAAATTTTAAATGAACAGGCAATTTGTGGCATCGCCCTATAGTCCACGCCACTCAAG	2030
Db	7246	AATACAA-----AATTAGCCAGGCATGCTGSCACATGCCCTGTAATCCCACTACTGAG	7193
Qy	2031	AGGCACAGCGGGAGGATCAGTCTGAGCCTGGGAGTGTGTTGTCAGTGACCTATGATTG	2090
Db	7192	AGGCTGAGCGAGGAGATCACTTGAACCGGGAGCGGAGGTTGCCGTGAGGTGAGATCG	7133
Qy	2091	TACACTGCACTCAGCCTGGGCAAC--AGAGCAACACCTTGCTCAAAAAATAAAACAACT	2149
Db	7132	CGCCTGCACTCCAGCCTGGGCAACAGAGCGGAACCTCCGCTCTCAAAAAAATAAA	7075
Qy	2150	AAATTTAAAAAAGAGACAGAGATAGTGGGTGTGGTGCACACCTGCAATCCCAGC	2209
Db	7075	AAAGTATATATTTTAAAGACATTGGCGGTGGCTCAGCCCTGTAAATCCCAGC	7016
Qy	2210	ACTTTGAGGCGGAGTGGGAGATCATCTGAGCCAGGAGTTTCAAGACCAGCCTGGCT	2259
Db	7015	ACTTTGGGAGCGGAGTGGGAGATCA--CGAGCTCAGGAGATCGAGACCATCCTGGCC	6958
Qy	2270	AACATGTTGAATTCCTATCTCTACCAAAAATACAAAATTAGCCAGGGCTGCTGGTGGC	2329
Db	6957	ACACAGTAAMACCCCGTCTCTACTAAAAATACAAAATTAGCTGGGCACGGTGGTGCAT	6898
Qy	2330	ACCTGACTGGGAGGTGCCCAACCCAGCTACTGGGGAGGCTGAGTCAGGAGAATCGCTTG	2389
Db	6897	GCCTGTA-----AACCCAGCTACCAAGGTACTCGGGAGGCTGAGGCAGGACAATCGCTTG	6844
Qy	2390	AACCTGGGAGCGGAGTTGGGTCAGCTGAGATGGTGCCACTGCACCTCGACCTCGGGC	2449
Db	6843	AACCGGAGTCGGAGGTTTGGGTGAGCTGAGATCATGCCACTGCACCTCGCGGCTGGAGA	6784
Qy	2450	AAAGAGCGCACTCTGTCTCCAAAAAAGAGAGAGGAGGACGACACAGA	2497
Db	6783	CAAGACAGACTCTCCCTCTCAAAAAAATAAAATAAAAGAAAAAATAAAAAA	6736
RESULT	9		
T94101/C			
ID	T94101	standard; DNA; 53526 BP.	
AC	T94101;		
DE	01-JUN-1998	(first entry)	
DE	Human PKD1 gene.		
KW	Human; polycystic kidney disease 1; PKD1; treatment;		
KW	autosomal dominant polycystic kidney disease; APKD; ss.		
OS	Homo sapiens.		
PN	W09744457-AL.		
PD	27-NOV-1997.		
PF	22-MAY-1997; U08799.		
PR	03-JUN-1996; US-658136.		
PR	24-MAY-1996; US-655360.		
PA	(GENZ.) GENZYME CORP.		
PI	Burn T, Connors T, Dackowski W, Germino G, Klinger K,		
PI	Qian F;		
DR	WPI; 98-018511/02		

PT	Human polycystic kidney disease gene, PKD1 - useful to treat and
PS	diagnose human autosomal or adult onset polycystic kidney disease
PP	Claim 2; Pages 90-118; 257pp; English.
CC	The present sequence is the human polycystic kidney disease 1
CC	(PKD1) gene. The PKD1 gene or polypeptide may be used to treat
CC	autosomal dominant polycystic kidney disease (APKD), and identify
CC	carriers of mutant PKD1 genes, i.e. subjects susceptible to APKD.
CC	Antibodies (Ab) that distinguish between normal and mutant PKD1
CC	sequences can also be used in diagnostic tests. Anti-PKD1 Ab may
CC	also be used to perform subcellular and histochemical localisation
CC	studies, and to block the function of PKD1. Ab are also useful in
CC	rational drug design studies to identify and test inhibitors of
CC	PKD1. Sense and antisense sequences derived from the PKD1 gene may
CC	be used for detection and therapy.
SQ	Sequence 53526 Bp; 8486 A. 17665 C; 15768 G; 11607 T

Query Match 7.8%; Score 381.2; DB 1; Length 53526;
Best Local Similarity 67.3%; Pred. No. 4.6e-42;
Matches 719; Conservative 0; Mismatches 228; Indels 121; Gaps 8;

1534	QY	CTCAGCCCTGTTAATCCAGCACTTTTAAGAGGCCAAGGCGGATGGATCACATTGAGCCCAAGG	1593
7787	Db	CTCAGACCTGTAAATCCAGCACTTTGGGAGCCGAGCAAGCATCACCTAAAGGCCAAG	7728
1594	QY	AGTTCAACACCAAGCTGAGCAACATGGTAAACCCCACTCTCTACCAAAATA- ---CAA	1649
7727	Db	AGTTCAAGACCAAGCTGACCAACATGGTGAACCCCACTCTCTACTAAAAATATTTTAAA	7668
1650	QY	AATTAGCTGGGCTTGGTGGCTGGCCCTGTAAATCCAGCTACTCAGGAGACTCAGGCAGA	1709
7667	Db	AATTAGCGCGGTGTGGCGCGCCCTCTAATCCAGCTACTCAGGAGCTCAGGCAGG	7608
1710	QY	AGATCGCTTGAACCTGGAGCGCAGAGTTGAGTGAGCGAGATTTGCGGCCACTGGAATC	1769
7607	Db	AGAAATCGCTTGAACCCGGGAGATGAGGAGCTGCAGCTGAGCCGAGATCACCAATTTGCCTC	7548
1770	QY	CAGCTGCGCTGACAGAGAGCCTGTCTCTAAATTAATTAATTAATTAATTAATTAATCA- ---	1826
7547	Db	CAGCCTGGGTGACAGAGCCAGACTCCGTCTCAACAAACAAACAAACAAAGACATCAGCTA	7488
1826	QY	-----	1826
7487	Db	GCTGTGCCAAGCACAGTGGTGTTCACACGAATTGATCACAGCCAGGTAGAAATTCCTTCAT	7428
1826	QY	-----	1851
7427	Db	TCTTTCTCCAGTCTACTGCTTTGCTTGACCAGCCTTAAAGACACACATATATTTTGTG	7368
1852	QY	CAGGTGCACTGGCTCAGGCCTAATATCTCAGCACTCTGGGAGGCCAA- GATGAGGAGTTG	1910
7367	Db	TGGGCGGTTTGCTCACACCTGTATCCCAACACTTTGGGAGGCCAAGCAGCGCGGATCA	7308
1911	QY	CTTGAAGCAGGAGTTTGGGACCAGCCTGGGCAACATAGGGGGGATCCCATCTCTTACACAC	1970
7307	Db	CCTGAGGTCAGGAGTTTGAGACACCGCTGACCAAGTGGAGAAACCCGCTCTCTCCTAAA	7248
1971	QY	AAAAAAATTTTAAATGAACAGGCAATTGTGGCATTGCOCCTATAGTCCGACCACTCAAG	2030
7247	Db	AATACAA- ----AATTAGCCAGGATGGTGACATCCCTGTAAATCCCAAGCTACTGGAG	7194
2031	QY	AGGCACAGCGGGAGGATCACATTGAGCTGGGAGTTCTGCTGTCAGTGAGCTATGATTG	2090
7193	Db	AGGCTGAGCGAGGAGTAATCATTGAACCGGGAGCGGAGGTTGCCGTGAGGTGAGATCG	7134
2091	QY	TAGCACTGCATCCAGCCTTGGGCAAC- AGAGCAAGACCTTGTCTCAAAAATAAACAAACT	2149
7133	Db	CGCCACTGCATCCAGCCTTGGGCAACAGAGCGAACTCCCTCTCAAAAAA- ---	7076
2150	QY	AAATTAATAAAGAGACGAGAGATAGTGGGTGTGGTGCCTCACACCTTGCATTCACAGC	2209
7076	Db	AAAGTATATATTTTAAAGACATTTGGCGGGTGGCGTGCAGGCTGTAAATCCCAAGC	7017

Qy	2210	ACTTTGGAGGCGCGAGGTGGGCGAGATCATCTGTAGGCCACGAGAGTTTCAGACACCGCTGGCT	2269
Db	7016	ACTTTGGGAGGCGCGAGGTGGGCGAGATCA--CGAGGTCTAGGAGATCGAGACCATCTCGGCC	6959
Qy	2270	AACATGTCGAATTCCTATCTCTACCAAAAATACAAAAATTAGCCAGCGCTGGTGGGCG	2329
Db	6958	AACACAGATAAACCCCGTCTCTACTATAAATAACAAAAATTAGCTGGGCACCGTGGTGCAT	6899
Qy	2330	ACCTGTACTGGGAGGTGCCACCCAGCTACTTGGGGAGGCTGAGTCAGGAGAATCGCTTG	2389
Db	6898	GCCTGTA-----AACCCAGCTACCAGGTACTCTGGGAGGCTGAGGAGGAGAATCGCTTG	6845
Qy	2390	AACCTGGGAGGCGGAGGTTGCGGTGAGCTGAGATGGTGCCACTGCATCCAGCTCGGCGG	2449
Db	6844	AACCCAGGAGTGGGAGTTGCGGTGAGCTGAGATCATGCCACTGCATCGCGCTGGAGA	6785
Qy	2450	AAAGACGCACTGTGCTCCAAAAAAGAGAGGAGGAGACACACAGA	2497
Db	6784	CAAGACCAAGACTCCGCTCTCAAAAAAATAAAAAAAGAAAAAATAAAAAA	6737
RESULT 10			
T78853/G			
ID	T78853 standard; DNA; 6901 BP.		
AC	T78853;		
DT	06-FEB-1998 (first entry)		
DE	Human lecithin-cholesterol acyltransferase (LCAT) DNA.		
KW	Human lecithin-cholesterol acyltransferase; LCAT; lecithin; cholestero		
KW	Artherosclerosis; heart disease; stroke; heart attack;		
KW	cholesterol reduction; peripheral vascular disease; Fish Eye Syndrome;		
OS	myocardial infarction; Classic LCAT Deficiency Syndrome; ss.		
KS	Homo sapiens.		
FT	Key	Location/Qualifiers	
FT	Intron	1..812	
FT		/*tag= a	
FT		/number= 1	
FT		/note= "intron start not specified"	
FT	misc_feature	709..716	
FT		/*tag= b	
FT		/note= "region indicated in figure 5, but no other	
FT		information provided"	
FT	misc_feature	786..790	
FT		/*tag= c	
FT		/note= "region indicated on figure 5, but no other	
FT		information provided"	
FT	misc_feature	809	
FT		/*tag= d	
FT		/note= "position is marked, but no further information	
FT		is given"	
FT	exon	813..990	
FT		/*tag= e	
FT		/number= 1	
FT	CDS	837..5092	
FT		/*tag= f	
FT		/note= "includes introns"	
FT	sig_peptide	837..902	
FT		/*tag= g	
FT	mat_peptide	909..5029	
FT		/*tag= h	
FT		/note= "introns included in sequence"	
FT	Intron	991..1723	
FT		/*tag= i	
FT		/number= 2	
FT	exon	1724..1880	
FT		/*tag= j	
FT		/number= 2	
FT	Intron	1881..1959	
FT		/*tag= k	
FT		/number= 3	
FT	exon	1960..2075	
FT		/*tag= l	
FT		/number= 3	
FT	Intron	2076..2169	

RESULT 10

T/8853/C
ID T78853 standard: DNA: 6901 BP.

AC T78853;
DT 06-FEB-1998 (first entry)

Human lecithin-cholesterol acyltransferase (LCAT) DNA.

KW Arteriosclerosis; heart disease; stroke; heart attack

KW cholesterol reduction; peripheral vascular disease; F

KW myocardial infarction; classic LCAT deficiency syndrome

OS	Homo sapiens.	Location/Qualifiers
FH	Key	

FT	intron	1. .812
FT		

FT /number= 1

```

FI      /note= "Intrusion start not specified"
ET      misc_feature 709. .716

```

```

FT      b
FT      /*tag=
FT      /note= "region indicated in figure 5.

```

FT		Information provided"	
FM	- 706 -	- 706 -	706

```

ET
m30c_2000000
/*tag= c
00.170

```

ET
ET

/note= "Region indicated on figure 3,
information provided"

```
FT misc_feature 809 /**acm d
```

FT /note= "position is marked, but no fur

FT	FT exon	813, .990	is given
1			
2			
3			
4			
5			
6			
7			
8			
9			
10			
11			
12			
13			
14			
15			
16			
17			
18			
19			
20			
21			
22			
23			
24			
25			
26			
27			
28			
29			
30			
31			
32			
33			
34			
35			
36			
37			
38			
39			
40			
41			
42			
43			
44			
45			
46			
47			
48			
49			
50			
51			
52			
53			
54			
55			
56			
57			
58			
59			
60			
61			
62			
63			
64			
65			
66			
67			
68			
69			
70			
71			
72			
73			
74			
75			
76			
77			
78			
79			
80			
81			
82			
83			
84			
85			
86			
87			
88			
89			
90			
91			
92			
93			
94			
95			
96			
97			
98			
99			
100			

```
ET /tag= e
ET /number= 1
```

```
FT CDS
FT 837.5092
FT /tag= f
```

FT	note= "includes introns"
027	003

```

FT
  orig=reference
  conf=1.000
  /*tag= g

```

```
FT mat_peptide . 909 . 5029
FT /*tag= h
```

FT	/note="introns included in sequence"
FT	991 1723
	{ intron

```

FT
...
/*tag= 1
...

```

FT	exon	1724..1880	transcript
11			

```

ET /tag= }
ET /number= 2

```

ET
intron
1881. .1959
/4+acm y

ET
/number= 3

```

FT
E1
EXON
1000: .2073
/*tag= 1

```

FT	number= 3
FT	2076. 2169
intron	

RESULT 11

V35620
 ID V35620 standard; DNA; 32367 BP.
 AC V35620;
 DT 07-SEP-1998 (first entry)
 DE Human SHOX (short stature homeobox containing gene) gene sequence.
 KW Homeobox domain; human growth gene; growth regulation; growth defect;
 KW Turner's syndrome; short stature homeobox containing gene; SHOXa;
 KW SHOX; bone disease; osteoporosis; calcium regulation; short stature;
 KW transcription factor A; ss.
 OS Homo sapiens.
 PN WO9814568-A1.
 PD 09-APR-1998.
 PF 29-SEP-1997; E05355.
 PR 16-JAN-1997; EP-100583.
 PR 01-OCT-1996; US-027633.
 PA (RAPP/) RAPPOLO-HOERBRAND G.
 PI Rao E. Rappold-hoerbrand G;
 DR WPI; 98-271719/24.
 PT New human growth genes - used to develop products for the diagnosis
 PT and treatment of human growth defects such as short stature, e.g.
 PT Turner's syndrome
 PS Claim 19; Pages 51-67; 84pp; English.
 CC This is the human SHOX gene sequence containing the PARI region. The gene
 CC region corresponding to short stature has been identified as a region of
 CC approximately 500 kb in the PARI region of the X and Y chromosomes. Three
 CC genes in this region have been identified as candidates for the short
 CC stature gene. These genes were designated SHOX (also referred to as
 CC SHOX93 or HOX93), pET92 and SHOT (SHOX-like homeobox gene on chromosome
 CC three). The SHOX gene has two separate splicing sites resulting in two
 CC variations SHOXa and SHOXb. The specification provides sequences of SHOX
 CC (short stature homeobox-containing) genes SHOX ET92, SHOXa, SHOXb, SHOT
 CC and exons of the SHOX genes as shown in V35610 to V35621 and protein
 CC sequences of the human growth protein transcription factor SHOXa, SHOXb
 CC and SHOT as shown W60573 to W60575. The novel genes are responsible for
 CC human growth. Defects in the genes can cause short stature, e.g.
 CC Turner's syndrome. The products can be used to develop agents for the
 CC treatment of short stature or other human growth disorders. The products
 CC can also be used for providing a mitogenic effect on cells, e.g. for the
 CC treatment of bone diseases such as osteoporosis and diseases involved
 CC with disturbance in the bone calcium regulation.
 SQ Sequence 32367 BP; 7627 A; 8130 C; 8564 G; 8043 T;

QY 1864 CTCAGCGCTATATCTCAGCACTCTGGAGG-CCAAGATGGAGGATGCTTGAAGCCAGG 1922
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 10484 CTCAGCGCTGTATCCAGCACTTTGGAGGCCAGGAGGTGGATGCTTGAAGTCAGG 10543
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 QY 1923 AGTTGGGACCAAGCTGGGCAACATAGGGGATCCCATCTCTACACACAAAAATTTT 1982
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 10544 AGTTGGAGACCAAGCTGGGCAACATAGTGAACCCCATCTCTATAAATAACAACGTT 10603
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 QY 1983 TAATGAACCAAGCAATTTGGCATCGGCTATAGTCCAGCACTCAAGAGGACAGGCGG 2042
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 10604 -----AGCTGGGTGTGGTGTGGCTGTATATCCAGCTACTCGGGAAGCTGAGGCTG 10658
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 QY 2043 GAGGATCACTTGAGCCCTGGGAGGTTGTGGTGCAGTGAGCTATATGTACCACCTGCACT 2102
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 10659 GAGAATTTGCTTGAATCTGGGAGGTGGAGGTTGCAGGAGGCGAGATAGTGCACCTGCACT 10718
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 QY 2103 CCAGCTGGGCAACAGAGCAAGACCTTGTCTCAAAAATAAACAACATAAAATTAATAAAAA 2162
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 10719 CCAGCTGGGCAACAGAGCAAGACCTCGCTCTCAAAAACAAGAAAGCAAAAAA 10778
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 QY 2163 GAAGACGAGAGATAGTGGGTGTGGTGTCTACACCTTCCAGCACTTTTGAAGGCC 2222
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 10779 CAAG----- 10783
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 QY 2223 GAGTGGGCAGATCATCTGAGGCCAGGAGTTCAAGACAGCAGCTGGCTAACATGTGTAAT 2282
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 10783 -----AGACCAGCTGGCCCAACATGTGTAAC 10809
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 QY 2283 CCTATCTACAAAATAACAAAATAGCCAGCGGTGGTGGGCACTGTACTGGGG 2342
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 10810 CGGTCTCTACTAATAA--TACAAATTAGCCGGCATGGTGGGCACTGTAGT----- 10864
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 QY 2343 AGGTGCCCAACAGCACTACTGGGGAGGCTGAGTCAGGAGAATCGCTTGAACCTGGGAGCG 2402
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 10864 -----CCAGCTACTCGGAGGCTGAGCGAGGAGAATGGCTTGAACCTGGGAGTG 10914
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 QY 2403 GAGTGGGTGCTGAGATGGTGCACCTGCATCTCAGCTGGGCGAAGAGGAGCTCT 2462
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 10915 GAGCTTGCACTGAGCGAGAGATAGTGCACCTGCATCTCCAGCTGGGCGACAGAGGAGCT 10974
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 QY 2463 GTCTCCAAAAAAGAGAGAGGAGGAGGACACAGACACACAGAGAGAAAGCC 2517
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 10975 TGATTTCAAGAACCCACCACCAACAAAAAACAACAAAAATCCAAAAAACC 11029
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 12

V35616
 ID V35616 standard; DNA; 15577 BP.
 AC V35616;
 DT 07-SEP-1998 (first entry)
 DE SHOX gene preliminary nucleotide sequence (HOX93).
 KW Homeobox domain; human growth gene; growth regulation; growth defect;
 KW Turner's syndrome; short stature homeobox containing gene; short stature;
 KW SHOX; bone disease; osteoporosis; calcium regulation; HOX93; ss.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT exon 1498..1807
 FT /*tag= a
 FT /note= "part of exon I (G310)"
 FT misc_feature 3844..4088
 FT /*tag= b
 FT /note= "pET92 region (first part)"
 FT misc_feature 4326..4437
 FT /*tag= c
 FT /note= "pET92 region (second part)"
 FT misc_feature 4545..4619
 FT /*tag= d
 FT /note= "pET92 region (third part)"
 FT exon 5305..5512
 FT /*tag= e
 FT /note= "part of exon II (ET93)"
 FT exon 11620..11729

FT WO9814568-A1.
PD 09-APR-1998.
PN 29-SEP-1997; E05355.
PF 16-JAN-1997; EP-100583.
PR 01-OCT-1996; US-027633.
PA (RAPF/) RAPPOLD-HOERBRAND G.
PI Rao E, Rappold-Hoerbrand G;
DR WPI: 98-271719/24.
PT New human growth genes - used to develop products for the diagnosis
PT and treatment of human growth defects such as short stature, e.g.
PT Turner's syndrome
PS Disclosure; Pages 37-45; 84pp; English.
CC This is a preliminary nucleotide sequence of the SHOX gene. The gene
CC region corresponding to short stature has been identified as a region of
CC approximately 500 kb in the PAR1 region of the X and Y chromosomes. Three
CC genes in this region have been identified as candidates for the short
CC stature gene. These genes were designated SHOX (also referred to as
CC SHOX93 or HOX93), PEX2 and SHOT (SHOX-like homeobox gene on chromosome
CC three). The SHOX gene has two separate splicing sites resulting in two
CC variations SHOXa and SHOXb. The specification provides sequences of SHOX
CC (short stature homeobox-containing) genes SHOX E92, SHOXa, SHOXb, SHOT
CC and exons of the SHOX genes as shown in V35610 to V35621 and protein
CC sequences of the human growth protein transcription factor SHOXa, SHOXb
CC and SHOT as shown W60573 to W60575. The novel genes are responsible for
CC human growth. Defects in the genes can cause short stature, e.g.
CC Turner's syndrome. The products can be used to develop agents for the
CC treatment of short stature or other human growth disorders. The products
CC can also be used for providing a mitogenic effect on cells, e.g. for the
CC treatment of bone diseases such as osteoporosis and diseases involved
CC with disturbance in the bone calcium regulation.
SQ Sequence 15577 BP; 3692 A; 3875 C; 4196 G; 3629 T;

/*tag= f
/note= 'part of exon IV (G108)'

Query Match 7.1%; Score 345.8; DB 1; Length 15577;
Best Local Similarity 66.8%; Pred. No. 1.9e-37;
Matches 678; Conservative 0; Mismatches 212; Indels 125; Gaps 8;
QY 1518 AGCCAGGCTGTGCTCAGCCTGTATCCAGCAGCTTATAGAGCCAGCGGATG 1577
DB 10335 AGTCAGGTGGGTGCTACACCTCTATCCAGCAGCTTTGGAGGCCAGGAGGTG 10394
QY 1578 ATCATGTAGCCAGGAGTTCAACACAGCAGCTGTAGCAATGTGTAACCCCATCTCTAC 1637
DB 10395 ATCGCTTGGTCCAGGAGTTGAGACAGCCTGGCCAAACATGTGAACCCCGTCTA- 10454
QY 1638 CAAAAATACAAAATTAAGTGGCCTGTGGTGGCTGTATCCAGCTACTCAGGA 1697
DB 10454 CTAAATACGAAATTAAGCCAGCATGGTGTGAGCACCTGTATTCAGCTACCGGA 10513
QY 1698 GACTGAGGAGAGGATCGCTTCAACCTGGAGGAGAGGTTGTCAGTGAGCGAGATTG 1757
DB 10514 GGCTGAGGAGAGATCGCTTGAACCTGGAGGTGACCTGTGATGAGTGAGATCAC 10573
QY 1758 GCACCTGGACTCCAGCCTCGGTGACAG- ---AGAGCCTGTCTCTAAATTAATTAAT 1813
DB 10574 ACCACTGCACTCCAGCCTGGATGAGAGAGCAAGACTGTCTCAAAAACAAATAAACA 10633
QY 1814 TANTTTAATTCATTTTAAAGACGAA- -----AAGTACGCGCAGGTGCGAGTG 1863
DB 10634 AAAAAACAAACAAATCAAAAAAGAAACCAATTTCCAGTTCTAGGCCAGGTGCGAGTG 10693
QY 1864 CTCACGCTTATATCTCAGCAGCTCTGGAGG-CCAGATGGAGGATGTCTTGAAGCCAGG 1922
DB 10694 CTCACGCTGTATCCAGCAGCTTTGGAGGCCAGAGGCTGATGCTGTAGGTGAGG 10753
QY 1923 AGTTTGGGACAGCCTGGGCAACATAGGGGGATCCCATCTCTACACACAAAAATTTT 1982
DB 10754 AGTTGAGACAGCCTGGGCAACATAGTGAACCCCATCTTTTACTAAAAATACAAAGT 10813
QY 1983 TAATGAACAGGATTTGTGGCATGGCCTATAGTCCAGCCATCAAGAGGACAGCGG 2042

DB 10814 -----AGCTGGGTGTGGTGTGCGCCTGTAAATCCAGCTACTCGGGAAGCTGAGGCTG 10868
QY 2043 GAGGATCACTTACAGCTGGGAGGTTGTGGTTCAGCTAGCTATGATTACCACTGCACT 2102
DB 10869 GAGAAATGCTTGAATCTGGGAGGTGGAGGTTCAGGGAGGGGAGATAGTGCACATGCACT 10928
QY 2103 CCAGCCTGGGCAACAGACAGCAAGACCTTGTCTCAAAAAATAAACAACTAAATAAAAAA 2162
DB 10929 CCAGCCTGGACAGAGACAGACTCCGCTCTCAAAAAACAAGAAGCAAAACAAAAA 10988
QY 2163 GAAGACGAGAGATAGTGGTGTGGTGTCAACACCTGCAATCCAGCACTTTTGGAGGCC 2222
DB 10989 CAAG----- 10993
QY 2223 GAGGTGGGAGATCATCTGAGGCCAGGAGTTCAAGCAGCAGCTGCTACATGGTGAAT 2282
DB 10993 -----AGACCAGCCTGGCCACACATGGTGAAT 11019
QY 2283 CCTATCTCTACCAAAAAATACAAAAATTAGCCAGGCGTGTGGTGGGCACCTGTACTGGG 2342
DB 11020 CCGCTCTTTACTAAAA--TACAAATTAGCCGGGATGTTGGTGGGCACCTGTACT- 11074
QY 2343 AGTGCCCAACAGCAGTCTGAGGAGGCTGAGTCAGAGAGATCGCTTGAACCTGGAGGG 2402
DB 11074 -----CCAGCTACTCTGGGAGGCTGAGGAGAGATGGCTTGAACCTGGAGGTG 11124
QY 2403 GAGGTGGGCTCAGCTGAGATGCTGCCACTGCACCTCCAGCCTGGCGGAAAGACACTCT 2462
DB 11125 GAGCTTGCAGTGGCGGAGATAGTCCACTTGCACCTGGGCGGACAGAGCGAGCT 11184
QY 2463 GTCTCAAAAAAAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2517
DB 11185 TGATTTGAGAACCCACCACCACCACCAACAAACAAACAAACAAACAAACAAAC 11239

RESULT 13
T75284
ID T75284 standard; DNA; 5543 BP.
AC T75284;
DT 21-DEC-1998 (first entry)
DE Nucleotide sequence encoding human RAD50 intron 19, 3' end.
KW ds; human; RAD50; DNA repair; tumour suppression; cancer; Septin-2;
KW central nervous system.
OS Homo sapiens.
PN WO9727284-A2.
PD 31-JUL-1997.
PF 24-JAN-1997; U01299.
PR 17-JUL-1996; US-687080.
PR 26-JAN-1996; US-592126.
PA (GENE-) GENELABS TECHNOLOGIES INC.
PI Dolganov G;
DR WPI: 97-393672/36.
PT Human tumour suppressor gene RAD50 - useful to detect
PT predisposition to, decrease risk of and treat cancer, also Septin-2
PT homologues
PS Claim 1; Page 122-124; 195pp; English.
CC The human RAD50 (hRAD50) is involved in DNA repair and has tumour
CC suppression activity, can be used to detect predisposition to, decrease
CC the risk of or treat cancers, e.g. acute myeloid leukaemia,
CC myelodysplastic syndrome, therapy related myelodysplastic syndrome,
CC therapy related acute myeloid leukaemia, refractory anaemia or refractory
CC anaemia with excess blasts. Also disclosed in this invention is Human
CC Septin-2 homologues of which may be used as targets for cancer therapies
CC and central nervous system directed treatment methods, and to measure the
CC proliferative potential of selected cell types.
SQ Sequence 5543 BP; 1718 A; 1044 C; 1110 G; 1671 T;

Query Match 6.9%; Score 339.4; DB 1; Length 5543;
Best Local Similarity 73.8%; Pred. No. 1.4e-36;
Matches 492; Conservative 0; Mismatches 156; Indels 19; Gaps 4;
QY 1499 GTCTCATAGATAGGAGAGGCCAGGCGCTGTGCTCAGCGCTGTATTCAGGACACTTT 1558

```
Db 630 GTAGTCAATTAAGAGGAGGAGCCAGGCGAGTCACTATCCCTGTATCCACCACTTT 689
Qy 1559 AAGAGGCAAGCGGATGATCACTTGAAGCCAGGAGTCAACACCACTGAGCAACAT 1618
Db 690 GGGAGGCAAGCGGATGATCACTTGAAGCCAGGAGTCAACACCACTGAGCAACAT 749
Qy 1619 GGTAAACCCATCTTACCAAAATACAAATAAGTCTGGCTGGTGGTGGCGCTG 1678
Db 750 GGTAAACCCATCTTACCAAAATAGAAAATAGAACAGTCTGGTGGTGGCGCTG 809
Qy 1679 TAAATCCAGCTACTCAGGAGACTGAGGAGAGGATGCTTGAACCTGGGAGGAGGAT 1738
Db 810 TAAATCCAGCTACTCAGGAGCTGAGGAGAGGATGCTTGAACCTGGGAGGAGGAT 869
Qy 1739 TGCAGTGGCGGAGATGCGCCACTGCACTCCAGCTCGCTGACAGAGGCTCTCTTA 1798
Db 870 TGCAGTGGCGGAGATGCACTCCAGCTCGCTGACAGAGGCTCTCTTA 928
Qy 1799 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1858
Db 929 TCTCAAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAAC 974
Qy 1859 AGTGGCTCACGCTATATCTCAGCACTCTGGGAGGCGCAAGATGGAGGATGCTTGAAGC 1918
Db 975 GGTGGCTCACGCTGTATATCCAGCACTTGGGAGGCGCAAGATGGAGGATGCTTGAAGC 1033
Qy 1919 CAGGAGTTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1978
Db 1034 CAGGAGTTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1090
Qy 1979 TTTTATGACACGAGGATGCTGAGCTGCGCTATATGCTGAGTATGCTGAGTATGCTGAG 2038
Db 1091 CAAATAATGACACGAGGATGCTGAGCTGCGCTATATGCTGAGTATGCTGAGTATGCTGAG 1150
Qy 2039 CGGGAGGAGTCACTTGAAGCTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2098
Db 1151 CGAGGAGTATGCTGAGCTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1210
Qy 2099 CACTCCAGCTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2158
Db 1211 CACTCCAGCTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1270
Qy 2159 AAAGAA 2165
Db 1271 GTCACAA 1277
```

RESULT 14

```
Q75209 ID Q75209 standard; cDNA; 8342 BP.
AC Q75209;
DT 23-AUG-1995 (first entry)
DE ALL-1 (acute lymphocytic leukaemia-1) breakpoint cluster region.
KW Acute lymphoblastic leukaemia; acute nonlymphoblastic leukaemia;
KW chromosomal translocation; abnormality; detection; rearrangement;
KW breakpoint cluster region; Alu repeat; chromosome 11; probe B859; ds.
OS Homo sapiens.
FH Key
FT exon Location/Qualifiers
FT 1. .263
FT /*tag= a
FT /number= 5
FT intron 264. .2352
FT /*tag= b
FT exon 593. .666
FT /*tag= c
FT /number= 6
FT intron 667. .798
FT /*tag= d
FT repeat_unit 799. .1108
FT /*tag= e
FT /rpt_type= OTHER
FT /note= "Alu repeat-a (Class J)"
```

```
FT repeat_unit 1119. .1420
FT /*tag= f
FT /rpt_type= OTHER
FT /note= "Alu repeat-b (Class Sx)"
FT repeat_unit 1432. .1716
FT /*tag= g
FT /rpt_type= OTHER
FT /note= "Alu repeat-c (Class Sb0)"
FT repeat_unit 1921. .2216
FT /*tag= h
FT /rpt_type= OTHER
FT /note= "Alu repeat-d (Class J)"
FT exon 2353. .2484
FT /*tag= i
FT /number= 7
FT intron 2485. .3031
FT /*tag= j
FT exon 3032. .3145
FT /*tag= k
FT /number= 8
FT intron 3146. .6787
FT /*tag= l
FT repeat_unit 3973. .4268
FT /*tag= m
FT /rpt_type= OTHER
FT /note= "Alu repeat-e (Class Sb0)"
FT repeat_unit 4764. .5094
FT /*tag= n
FT /rpt_type= OTHER
FT /note= "Alu repeat-f (Class J)"
FT repeat_unit 6072. .6362
FT /*tag= o
FT /rpt_type= OTHER
FT /note= "Alu repeat-g (Class S)"
FT exon 6788. .6934
FT /*tag= p
FT /number= 9
FT intron 6935. .7966
FT /note= "nucleotides 7429-7559 show about 80%
sequence identity to genomic sequences
found in 5' regulatory regions, 3' segments,
or in introns of several genes such as ApoA4,
Factor VIIIc subunit and G6PD"
FT repeat_unit 7164. .7427
FT /*tag= q
FT /rpt_type= OTHER
FT /*tag= r
FT /note= "Alu repeat-h (Class Sx)"
FT exon 7967. .8062
FT /*tag= s
FT /number= 10
FT intron 8063. .8303
FT /*tag= t
FT exon 8304. .8342
FT /*tag= u
FT /number= 11
FT W09426930-A.
FT 24-NOV-1994.
FT 22-APR-1994; U04496.
FT 14-MAY-1993; US-062443.
FT (UJJE-) UNIV JEFFERSON THOMAS.
FT Canaani E, Croce C;
FT WPI; 95-006818/01.
FT P-PSDB; R66467.
FT New acute lymphocytic leukaemia gene prods. - used for the
FT diagnosis and treatment of leukaemias, partic. acute
FT lymphoblastic or nonlymphoblastic leukaemia
FT Example 5; Fig 22; 207pp; English.
FT A phage clone, mg11.1, which spans the breakpoint cluster region in
FT the ALL-1 gene has been sequenced (Q75209). Eight Alu repeat
FT sequences were identified and classified based on criteria
FT published in Milosavljevic et al. (J.Mol.Evol. 32, 105-121, 1991).
FT The high concentration of Alu sequences within the area spanned by
```


Qy 1524 GGCCTGGTGGCTCAGCCTGTAAATCCAGCAGCTTTAAGAGGCCAAGGGGATGATCACT 1583
 Db 2547 AGAGAGTGGCTCAGCCTGTAAATCCAGCAGCTTTAAGAGGCCAAGGGGATGATCACT 2488
 Qy 1584 TGAGCCAGGAGTTCAACACAGCAGCTGAGCAATAGTAAACCCCATCTCTACCAAAA 1643
 Db 2487 TGAGTCAAGAGTTGAGACAGACCGGTGAGCATGTGAACCCCATCTTTACTAAAA 2428
 Qy 1644 TACAAAATATAGTGGGCTGGTGGCTGGGCGCTGTAAATCCAGCAGCTCTCAGGAGCTGA 1703
 Db 2427 TACAAAATATAGCGGGCATGGTGGCAGTGCCTGTAAATCCAGCAGCTCTCAGGAGCTGA 2368
 Qy 1704 GGCAGAGGATCGCTGAACTGGAGGCGAGAGGTTGAGTGGAGCCGAGATTCGCCCACT 1763
 Db 2367 GGCAGGAGATCACTTTAACTCCAGAGGAGAGGTTGAGTGGAGCTGAGATTCGCCCACT 2308
 Qy 1764 GGACTCCAGCCTGGTGCAGAGAGCCTGTCTCTAAATTAATTAATTAATTAATTAATTAAT 1823
 Db 2307 GCATCCAGCCTGGCAGACAGAGCTCACTCAAAA--AATAAATAAATAAATAAATAAAT 2250
 Qy 1824 CAATTTAAAGACGAAAGTGCAGGCGAGTGCAGTGGCTCAGCCTATATCTCAGC 1883
 Db 2249 AAAAAATAA-----TTTAGGCTAGACGAGTGGCTCATGCTGTAAATCCCAAGC 2201
 Qy 1884 ACTCTGGAGGCCAGAT-GGAGGATTCCTGAGGCCAGGAGTTGGAGCCAGCCTGGGC 1942
 Db 2200 ACTTTGGAGGCTGAGTGGGAGGATCATCTGAGGTGAGGATTTGAGCCAGCCTGGCC 2141
 Qy 1943 AACATAGGGGATCCCATCTCTACACACAAAAAATTTTTAAATGAACCGGCAATTTGG 2002
 Db 2140 AACATGTTGAACCTGTCTCCAC-----TAAAAATACAAAAATAGCAGGTGTGGTGG 2086
 Qy 2003 CATGGCCTATAGTCCAGCCACTCAAGAGCCAGGCGGAGGAGTCACTTTAGCCTGGG 2062
 Db 2085 TGCATGCTGTAAATCCAGTACTCAGGAGGCTCAGGAGGAGTGCCTTGAACCTGGG 2026
 Qy 2063 AGTTGTGGTTGCAGTGAAGTATGATGTACCACTGCAGCTCCAGCCTGGGCAAC-AGAGC 2121
 Db 2025 AGGTAGAGGTTGCAGTGGGCCAGATTCCTCTCAGTCACTCCATCTCCCTGGGCAAGAGT 1966
 Qy 2122 AAGACCTGTCTCAAAAATAAACAACCTAAAAATAAAAAAGAAGA 2167
 Db 1965 GAACTGTCTCAAAAATAAATAAATAAATAAATAAATAAATAAATA 1920

Search completed: October 29, 1999, 02:38:48
 Job time: 7875 sec

This Page Intentionally Left Blank

GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 31, 1999, 00:45:31 ; Search time 1973.24 Seconds
(without alignments)
7889.367 Million cell updates/sec

Title: US-09-371-333-1
Perfect score: 4895
Sequence: 1 CCCCCACGGCTGGTGGCA.....AAAAAAAAAAAAAAAAAAAA 4895

Scoring table: IDENTITY_NUC

Searched: 679419 seqs, 1590154680 residues

Database : GenEmbl.*

- 1: gb_bal.*
- 2: gb_ba2.*
- 3: gb_cm.*
- 4: gb_ov.*
- 5: gb_pat.*
- 6: gb_ph.*
- 7: gb_pl1.*
- 8: gb_pl2.*
- 9: gb_pr1.*
- 10: gb_pr2.*
- 11: gb_pr3.*
- 12: gb_ro.*
- 13: gb_st.*
- 14: gb_sts.*
- 15: gb_sy.*
- 16: gb_un.*
- 17: gb_vi.*
- 18: em_fun.*
- 19: em_htg.*
- 20: em_hum1.*
- 21: em_hum2.*
- 22: em_in.*
- 23: em_om.*
- 24: em_or.*
- 25: em_ov.*
- 26: em_pat.*
- 27: em_ph.*
- 28: em_pl.*
- 29: em_ro.*
- 30: em_sts.*
- 31: em_sy.*
- 32: em_un.*
- 33: em_vi.*
- 34: gb_htg1.*
- 35: gb_htg2.*
- 36: gb_in1.*
- 37: gb_in2.*
- 38: em_bal.*
- 39: em_ba2.*
- 40: em_hum3.*
- 41: em_hum4.*
- 42: gb_pr4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Match	Length	DB	ID	Description
1	4895	100.0	4895	11	AF055917	Homo sapi	

2	1531.4	31.3	1534	11	AF080214	Homo sapi
3	681.6	13.9	1361	12	AF080215	Mus muscu
4	503.2	10.3	173616	35	AC006509	Homo sapi
5	494	10.1	45328	10	AC002128	Human DNA
6	494	10.1	41936	10	CH19R30879	Homo sapi
7	489.2	10.0	145414	9	HS78F24	Homo sapi
8	489.2	10.0	134506	11	AC005738	Homo sapi
9	489	10.0	42416	11	AC005551	Homo sapi
10	486	9.9	122320	10	HS216E10	Human DNA
11	484	9.9	188477	34	AC004971	Homo sapi
12	484	9.9	109613	34	HS550H1	Homo sapi
13	482.8	9.9	199827	42	AC007055	Homo sapi
14	482.2	9.9	44957	11	AC005372	Homo sapi
15	481.4	9.8	44826	9	HS426N21	Human DNA
16	481.2	9.8	152083	35	AC005399	Homo sapi
17	480	9.8	148008	34	HS27Q5	Homo sapi
18	477.8	9.8	16689	10	HS4011712	Homo sapi
19	477.4	9.8	158905	11	AC005666	Homo sapi
20	477	9.7	104848	34	AC004588	Homo sapi
21	473.6	9.7	35587	11	AC005763	Homo sapi
22	473.6	9.7	45858	11	AC006123	Homo sapi
23	473.2	9.7	108523	42	AC000159	Homo sapi
24	473	9.7	40946	11	AC002984	Human DNA
25	472.8	9.7	109613	34	HS550H1	Homo sapi
26	471	9.6	133478	11	AC004686	Homo sapi
27	468.4	9.6	196439	34	AC005867	Homo sapi
28	467.2	9.5	80023	34	HS359F1	Homo sapi
29	467.2	9.5	169254	34	HS892F13	Homo sapi
30	467.2	9.5	169254	34	HS892F13	Homo sapi
31	466.2	9.5	126322	11	AC000026	Homo sapi
32	466.2	9.5	173029	42	AC002059	Homo sapi
33	465.8	9.5	46353	34	HS439A6	Homo sapi
34	465.6	9.5	91880	34	HS633O19	Homo sapi
35	465.4	9.5	118899	10	HS1085F17	Human DNA
36	464.6	9.5	80141	10	HSW5G8R	Homo sapien
37	464	9.5	106921	42	AC004263	Homo sapi
38	463.8	9.5	57434	11	AC005808	Homo sapi
39	463.6	9.5	190508	42	AC005037	Homo sapi
40	463.4	9.5	39838	42	AC007766	Homo sapi
41	462.8	9.5	128328	11	AC005839	Homo sapi
42	462.6	9.5	128117	11	AC004000	Human PAC
43	462.2	9.4	162485	34	AC004840	Homo sapi
44	462.2	9.4	224895	34	HSC198A12	Homo sapi
45	461.6	9.4	152083	35	AC005399	Homo sapi

ALIGNMENTS

RESULT 1	AF055917	4895 bp	08-JUL-1998
LOCUS	Homo sapiens protease-activated receptor 4 mRNA, complete cds.		
DEFINITION	AF055917		
ACCESSION	AF055917		
NID	3293321		
VERSION	AF055917.1	GI:3293321	
KEYWORDS	human.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	1		
AUTHORS	Xu, W., Andersen, H., Whitmore, T.E., Presnell, S.R., Yee, D.P., Ching, A., Gilbert, T., Davies, E.W., and Foster, D.C.		
TITLE	Cloning and characterization of human protease-activated receptor 4		
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 95 (12), 6642-6646 (1998)		
MEDLINE	98283984		
REFERENCE	2		
AUTHORS	Xu, W., Andersen, H., Whitmore, T.E., Gilbert, T., Davies, E.W., and Foster, D.C.		
TITLE	Direct Submission		
JOURNAL	Submitted (26-MAR-1998) Biochemistry, University of Washington, J Wing, NE Pacific Street, Seattle, WA 98195, USA		

FEATURES
source

Location/Qualifiers
1..4895
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="19"
/map="19p12"
176..1333
CDS
/note="PAR4; member of family of G-protein coupled
protease activated receptors"
/codon_start=1
/product="protease-activated receptor 4"
/protein_id="AAC25699.1"
/db_xref="pid:g3293322"
/db_xref="gi:3293322"
/translation="MWGRLLWPLVLFSLSGTQTPSYDESSTGGDDSTPSILP
APRGVGVGVANDSDTLPELPDSSRALLLGGWPTRLVPAALYGLVPLVGLPANGLLALWVL
ATQAPRLSTMLLNALADLALLALPRTIAYHLRGWFFGEACAKLATAALYGHM
YGSVLLAAVSLDRYLALVHLPRALRRLRGLMAAALPLTLQRTF
RLASDRVLCHDRLPLDAQASHWOPAFCTLLALGCFLLPMLLLCYKALPTLAASGR
RYGHALRUTAVVLASAVAFVPSNLLLLHYSDPSPSANGNLYGAYVPSLALSTLNSC
VDPFIYYVSAEFRDKVRAGLFRSPGDTVASKASAEGRSGMGTSHSLQ"
BASE COUNT 971 a 1573 c 1400 g 951 t
ORIGIN

Query Match 100.0%; Score 4895; DB 11; Length 4895;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 4895; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CTCCTACGGGCTGGTGGCAAGCGGCCCTGGTGGGTCTCGGGGGGAGGGGAGCCTTCC 60
DB 1 CTCCTACGGGCTGGTGGCAAGCGGCCCTGGTGGGTCTCGGGGGGAGGGGAGCCTTCC 60
QY 61 TGGTTTATCTCACGGCGCGATCTGCTCGCTCGCTCGGCTCGGCTCCAGAGCTGGGCTCAG 120
DB 61 TGGTTTATCTCACGGCGCGATCTGCTCGCTCGGCTCGGCTCCAGAGCTGGGCTCAG 120
QY 121 GGTCCGGGAGGAGCAAGCCTGAGGCCACAGCCAGCAGCAGCCTGAGTGCAGTGTG 180
DB 121 GGTCCGGGAGGAGCAAGCCTGAGGCCACAGCCAGCAGCAGCCTGAGTGCAGTGTG 180
QY 181 GGGGCGACTGCTTCTGGCCCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 240
DB 181 GGGGCGACTGCTTCTGGCCCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 240
QY 241 CCCCAGCGTCTACGACGAGCGGAGCACCAGGAGTGGTGGTGGTGGTGGTGGTGGTGGT 300
DB 241 CCCCAGCGTCTACGACGAGCGGAGCACCAGGAGTGGTGGTGGTGGTGGTGGTGGTGGT 300
QY 301 CTGCTGCCCCCGGGCTACCCAGGCAAGTCTGTGCCAATGACAGTGCAGCAGCCTTGA 360
DB 301 CTGCTGCCCCCGGGCTACCCAGGCAAGTCTGTGCCAATGACAGTGCAGTGCAGCAGCCTTGA 360
QY 361 GCTCCCGGACACTCACGGGCACTGCTTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGG 420
DB 361 GCTCCCGGACACTCACGGGCACTGCTTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGG 420
QY 421 CCTCTATGGGCTGGTCTGGTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGT 480
DB 421 CCTCTATGGGCTGGTCTGGTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGT 480
QY 481 GGCCAGGAGGACCTCGGCTGCCCTCCACATGCTGTGATGAACCTCGGACCTGTGA 540
DB 481 GGCCAGGAGGACCTCGGCTGCCCTCCACATGCTGTGATGAACCTCGGACCTGTGA 540
QY 541 CCTCTGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGG 600
DB 541 CCTCTGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGG 600
QY 601 GCCCTTCGGGAGGCGGCTGCGGCTGGCCACGCGGCACTCTATGTGTACATGTATGG 660
DB 601 GCCCTTCGGGAGGCGGCTGCGGCTGGCCACGCGGCACTCTATGTGTACATGTATGG 660

QY 661 CTCAGTGTCTGTGCTGGCGCCGTCAGCTGTGATCGCTACCTGGCCCTGGCTGTGCACCCGCT 720
DB 661 CTCAGTGTCTGTGCTGGCGCCGTCAGCTGTGATCGCTACCTGGCCCTGGCTGTGCACCCGCT 720
QY 721 GCGGGCCCGCCCTGCTGCTGGCGCCGCTGGCCCTTGGACTCTGTCATGGCTGTGGCT 780
DB 721 GCGGGCCCGCCCTGCTGCTGGCGCCGCTGGCCCTTGGACTCTGTCATGGCTGTGGCT 780
QY 781 CATGGCGCCCGCCCTGGGACACTGCCCTGACACTGAGGGGCGAGACCTTCCGGCTGGCGCG 840
DB 781 CATGGCGCCCGCCCTGGGACACTGCCCTGACACTGAGGGGCGAGACCTTCCGGCTGGCGCG 840
QY 841 CTCGGATCGCTGTCTGCCATGACGCGCTGCCCTGGACGACAGGGCTCCCACTTGCA 900
DB 841 CTCGGATCGCTGTCTGCCATGACGCGCTGCCCTGGACGACAGGGCTCCCACTTGCA 900
QY 901 ACCGGCTTACCTGCTGGCGCTGTTGGGCTGTTTCTGCTGGCTGTGGCGCATGCTGCT 960
DB 901 ACCGGCTTACCTGCTGGCGCTGTTGGGCTGTTTCTGCTGGCTGTGGCGCATGCTGCT 960
QY 961 GTGCTACGGGGCCACCTGTCACAGCTGGCGCCAGCGGGCGGCGCTACGGCCACGGGCT 1020
DB 961 GTGCTACGGGGCCACCTGTCACAGCTGGCGCCAGCGGGCGGCGCTACGGCCACGGGCT 1020
QY 1021 GAGGCTGACCGAGTGTGCTGGGCTCCGCGCTGGCCCTTCTTCTGTCGCCAGCAACCTGCT 1080
DB 1021 GAGGCTGACCGAGTGTGCTGGGCTCCGCGCTGGCCCTTCTTCTGTCGCCAGCAACCTGCT 1080
QY 1081 GCTGCTGTGATTAATCTGCGGACCGAGCCCGAGCCCTGCGGCAACCTCTATGTTGCTA 1140
DB 1081 GCTGCTGTGATTAATCTGCGGACCGAGCCCGAGCCCTGCGGCAACCTCTATGTTGCTA 1140
QY 1141 CGTGCCAGCTGCGCTGAGCACCTCAACAGCTGGGTGATCCCTTCATCTACTACTA 1200
DB 1141 CGTGCCAGCTGCGCTGAGCACCTCAACAGCTGGGTGATCCCTTCATCTACTACTA 1200
QY 1201 CGTGTGCGCCGAGTTACAGGACAAAGTTCGGGCGAGGGCTCTTCCAACGGTGCAGCGGGGA 1260
DB 1201 CGTGTGCGCCGAGTTACAGGACAAAGTTCGGGCGAGGGCTCTTCCAACGGTGCAGCGGGGA 1260
QY 1261 CACGCTGGCCCTCAAGGCTCTGCGGAAGGGGCGAGCGGGGCGATGGCAGCCACTCCTC 1320
DB 1261 CACGCTGGCCCTCAAGGCTCTGCGGAAGGGGCGAGCGGGGCGATGGCAGCCACTCCTC 1320
QY 1321 TTTGCTCCAGTGACACAAGTGGGAGGCTGTACTGGTGCAGACAGGGTCCCTTCCGCC 1380
DB 1321 TTTGCTCCAGTGACACAAGTGGGAGGCTGTACTGGTGCAGACAGGGTCCCTTCCGCC 1380
QY 1381 ACTTCAGTCTCTTCTGGGACCTCAGAAATGTACCTTATTTGGAATAGGGTTGTTACAA 1440
DB 1381 ACTTCAGTCTCTTCTGGGACCTCAGAAATGTACCTTATTTGGAATAGGGTTGTTACAA 1440
QY 1441 CTGTCACTAGCGGAGGTCACCTTTGGAGAGGGTGGGCTTACATCCAGTGTGGGTGCTG 1500
DB 1441 CTGTCACTAGCGGAGGTCACCTTTGGAGAGGGTGGGCTTACATCCAGTGTGGGTGCTG 1500
QY 1501 CCTCATAGATTAAGGAGAGGCGAGGCTGGTGGCTGACGCTGTAAATCCAGCACTTAA 1560
DB 1501 CCTCATAGATTAAGGAGAGGCGAGGCTGGTGGCTGACGCTGTAAATCCAGCACTTAA 1560
QY 1561 GAGGCCAAGGGCGGATGATCACTTTGAGCCCGAGGATTCACACAGCGCTGAGCAACATGG 1620
DB 1561 GAGGCCAAGGGCGGATGATCACTTTGAGCCCGAGGATTCACACAGCGCTGAGCAACATGG 1620
QY 1621 TAAACCCCATCTCTACCAAAATACAAAATAGCTGGGCTTGGTGGCTGGGCTGGTGA 1680
DB 1621 TAAACCCCATCTCTACCAAAATACAAAATAGCTGGGCTTGGTGGCTGGGCTGGTGA 1680
QY 1681 ATCCAGCTACTCAGGAGCTGAGGCGAGAGGATCGGTTGAACCTGGGAGGAGGAGTTG 1740
DB 1681 ATCCAGCTACTCAGGAGCTGAGGCGAGAGGATCGGTTGAACCTGGGAGGAGGAGTTG 1740
QY 1741 CAGTGACCGGAGATTGGCCACTGGACTCCAGCTGGCTGAGCAGAGAGCCTGTCTCTAAA 1800

QY	3961	AAGTCAGGCGCGACGGGTGAGGACCGGCGCGGGGTGCGGCGCGGGCGCTA	4020	human.	KEYWORDS
Db	3961		4020	Homo sapiens	SOURCE
QY	4021	CCGGGTCGTAGTGTGTACAGGAGACTGGGAGCGCGGACGCTCTGCCACACGAC	4080	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;	ORGANISM
Db	4021		4080	Eutheria; Primates; Catarrhini; Hominoidea; Homo.	
QY	4081	TCCCGGAGACGACGACCGACGCTCAGGACCGGCTGGGATCTGTGGGGCAGCG	4140	1 (bases 1 to 1534)	REFERENCE
Db	4081		4140	Kahn,M.L., Hammes,S.R., Botka,C. and Coughlin,S.R.	AUTHORS
QY	4141	GCGGGCGAGGCTCGACCCGGGCGCAGGAGCGCGGGCGCTGAGCTCAGGCCACGAACTG	4200	Gene and locus structure and chromosomal localization of the	TITLE
Db	4141		4200	protease-activated receptor gene family	
QY	4201	GCTGATTTCCAGGATACCCAGGACGGTGAACACAGAGAAACGTTATCTTTCTT	4260	J. Biol. Chem. (1998) In press	JOURNAL
Db	4201		4260	2 (bases 1 to 1534)	REFERENCE
QY	4261	TTTTTCTTTTACTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT	4320	Kahn,M.L., Zheng,Y.W., Huang,W., Bigornia,V., Zeng,D., Moff,S.,	AUTHORS
Db	4261		4320	Faresse,R.V., Tam,C. and Coughlin,S.R.	TITLE
QY	4321	AGGCTGAGTGCAGTGGCGTGTCTGGGCTCACTGCAAGCTCGGCCTCTCGGCTTCAAT	4380	A dual thrombin receptor system for platelet activation	JOURNAL
Db	4321		4380	Nature (1998) In press	AUTHORS
QY	4381	GATTCCTGCTCAGCTCCCAAGTAGCTGGGATACAGGCGCCACACCGACCCCTG	4440	3 (bases 1 to 1534)	REFERENCE
Db	4381		4440	Kahn,M.L. and Coughlin,S.R.	AUTHORS
QY	4441	CTAATTTTGTATTTTGTATCAGAGGAGTTCACCATGTTGGCCAGGCTGGTCTCCA	4500	Direct Submission	TITLE
Db	4441		4500	Submitted (24-JUL-1998) CVRI, UCSF, 513 Parnassus, San Francisco,	JOURNAL
QY	4501	ACTCTGCCCTCAAGTGATCCGGCTCGGTCCCTATTTTATTTTATTTTATTTTATTT	4560	CA 94143-1030, USA	FEATURES
Db	4501		4560	Location/Qualifiers	
QY	4561	CAGTGGGAAACGCTCAGGTGGCTCTGAAACACCACTCTTTTGTGTGTGTGACGC	4620	1..1534	source
Db	4561		4620	/organism="Homo sapiens"	
QY	4621	ATGGCTGAGCATGTGTGGGTGGAGTCAGCACATTCAGGATCTGTGCAATCATCACCCTC	4680	/db_xref="taxon:9606"	
Db	4621		4680	/chromosome="19"	
QY	4681	TGCTTAGTTACAGGACGGTTCTTTCTCCCAAGAAACCCCATCGCCATCAGCACTCA	4740	/map="19p12"	
Db	4681		4740	3..1160	
QY	4741	CTCCCACTCCCCAGCCCTGCAACCAAAATCTTTTCAACTCTACGGATTGGCTGT	4800	/note="G protein-coupled receptor; thrombin receptor;	CDS
Db	4741		4800	PAR4"	
QY	4801	TCGGGCAATTCATGTCAATGGAATCATGTACTGTGTGAAAAAATAAAAAA 4895	4860	/codon_start=1	
Db	4801		4860	/product="protease-activated receptor 4"	
QY	4861	AAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA	4920	/protein_id="AAC28860.1"	
Db	4861		4920	/db_xref="PID:g3396081"	
QY	4921	ATGGCTGAGCATGTGTGGGTGGAGTCAGCACATTCAGGATCTGTGCAATCATCACCCTC	4980	/db_xref="GI:3396081"	
Db	4921		4980	/translation="MWGRLLWPLVLFSLGGTQTPSVYDESGTGGDDSTPSILP	
QY	4981	ATGGCTGAGCATGTGTGGGTGGAGTCAGCACATTCAGGATCTGTGCAATCATCACCCTC	5040	APRPGQVCANDSDTLELPDPSRALLGWYTRLPALYGLVLYVGPANGALWL	
Db	4981		5040	ATGAPLPSTMLLMNATLALLALPPRIAYHLRGQWPFGEACRLATPAAALYGHM	
QY	5041	ATGGCTGAGCATGTGTGGGTGGAGTCAGCACATTCAGGATCTGTGCAATCATCACCCTC	5100	YGSVLLAAVSLDRYLALVHPLRARALRGRLALGCMAAWMAAALATPLTQRTF	
Db	5041		5100	RYLGRDVLCHDALPDAQASHWOPAFCTLLAGLCFLPLLAMLLCYGATLTLAASGR	
QY	5101	ATGGCTGAGCATGTGTGGGTGGAGTCAGCACATTCAGGATCTGTGCAATCATCACCCTC	5160	RLGHALRLTAVLASAVFAFPVSNLLLLHYSDPSPSAWNLGYAVYPSLALSTLNSC	
Db	5101		5160	VDPIYYIVSAEFKDKVRAGLFQSPGDTVASKASAEGRSGMGTSHSLQ"	
QY	5161	ATGGCTGAGCATGTGTGGGTGGAGTCAGCACATTCAGGATCTGTGCAATCATCACCCTC	5220	BASE COUNT	
Db	5161		5220	235 a 526 c 479 g 293 t 1 others	
QY	5221	ATGGCTGAGCATGTGTGGGTGGAGTCAGCACATTCAGGATCTGTGCAATCATCACCCTC	5280	ORIGIN	
Db	5221		5280		
QY	5281	ATGGCTGAGCATGTGTGGGTGGAGTCAGCACATTCAGGATCTGTGCAATCATCACCCTC	5340	Query Match	
Db	5281		5340	31.3%; Score 1531.4; DB 11; Length 1534;	
QY	5341	ATGGCTGAGCATGTGTGGGTGGAGTCAGCACATTCAGGATCTGTGCAATCATCACCCTC	5400	Best Local Similarity	
Db	5341		5400	99.9%; Pred. No. 3.2e-242;	
QY	5401	ATGGCTGAGCATGTGTGGGTGGAGTCAGCACATTCAGGATCTGTGCAATCATCACCCTC	5460	Matches 1532; Conservative	
Db	5401		5460	0; Mismatches	
QY	5461	ATGGCTGAGCATGTGTGGGTGGAGTCAGCACATTCAGGATCTGTGCAATCATCACCCTC	5520	2; Indels	
Db	5461		5520	0; Gaps	
QY	5521	ATGGCTGAGCATGTGTGGGTGGAGTCAGCACATTCAGGATCTGTGCAATCATCACCCTC	5580	0;	
Db	5521		5580		
QY	5581	ATGGCTGAGCATGTGTGGGTGGAGTCAGCACATTCAGGATCTGTGCAATCATCACCCTC	5640	QY 174 TCATGTGGGGGCGACTGCTCTGTGGCCCTGTGTGTGGGGTTCAGCCTGTCTGGCGCA	233
Db	5581		5640	Db 1 TCATGTGGGGGCGACTGCTCTGTGGCCCTGTGTGTGGGGTTCAGCCTGTCTGGCGCA	60
QY	5641	ATGGCTGAGCATGTGTGGGTGGAGTCAGCACATTCAGGATCTGTGCAATCATCACCCTC	5700	QY 234 CCAGACCCCCAGGCTGTACAGCAGAGAGCGGAGCACCAGGAGGTGTGTATGACAGCACGC	293
Db	5641		5700	Db 61 CCAGACCCCCAGGCTGTACAGCAGAGAGCGGAGCACCAGGAGGTGTGTATGACAGCACGC	120
QY	5701	ATGGCTGAGCATGTGTGGGTGGAGTCAGCACATTCAGGATCTGTGCAATCATCACCCTC	5760	QY 294 CCTCAATCTGCTGCCCGCCCGGCTACCCAGGCCAAGCTGTGCCAATGACAGTGACA	353
Db	5701		5760	Db 121 CCTCAATCTGCTGCCCGCCCGGCTACCCAGGCCAAGCTGTGCCAATGACAGTGACA	180
QY	5761	ATGGCTGAGCATGTGTGGGTGGAGTCAGCACATTCAGGATCTGTGCAATCATCACCCTC	5820	QY 354 CCCTGGAGCTCCCGACAGCTCACGGCAGCTGCTTCTGGGCTGGGTGCCACAGGCTGG	413
Db	5761		5820	Db 181 CCCTGGAGCTCCCGACAGCTCACGGCAGCTGCTTCTGGGCTGGGTGCCACAGGCTGG	240
QY	5821	ATGGCTGAGCATGTGTGGGTGGAGTCAGCACATTCAGGATCTGTGCAATCATCACCCTC	5880	QY 414 TCGCCGCCCTCTATGGGCTGGTCTGTGGGTGGGCTGCCCGGCCAATGGGCTGGGCTGT	473
Db	5821		5880	Db 241 TCGCCGCCCTCTATGGGCTGGTCTGTGGGTGGGCTGCCCGGCCAATGGGCTGGGCTGT	300
QY	5881	ATGGCTGAGCATGTGTGGGTGGAGTCAGCACATTCAGGATCTGTGCAATCATCACCCTC	5940	QY 474 GGTGCTGTGCCACGAGGAGGACCTCGGCTGCCCTCCACCATGCTGTGTATGAACCTCGCA	533
Db	5881		5940	Db 301 GGTGCTGTGCCACGAGGACCTCGGCTGCCCTCCACCATGCTGTGTATGAACCTCGCA	360
QY	5941	ATGGCTGAGCATGTGTGGGTGGAGTCAGCACATTCAGGATCTGTGCAATCATCACCCTC	6000		
Db	5941		6000		

Qy	534	CTGCTGACCTCTGCTGGCCCTTGCGGCTTGCCCGCGCGATCGCCTACCACTGCGTGGCC	593
Db	361	CTGCTGACCTCTGCTGGCCCTTGCGGCTTGCCCGCGCGATCGCCTACCACTGCGTGGCC	420
Qy	594	AGGCGTGGCCCTTCGGGGAGCGCCCTCGCGCTTGCCCGACGCGCGACTCTATGGTCA	653
Db	421	AGGCGTGGCCCTTCGGGGAGCGCGCTCGCGCTTGCCCGACGCGCGACTCTATGGTCA	480
Qy	654	TGATATGGCTCAGTGTGCTGGCGCGCGCTCAGCCTGGATCGCTACCTGGCCCTGGTGC	713
Db	481	TGATATGGCTCAGTGTGCTGGCGCGCGCTCAGCCTGGATCGCTACCTGGCCCTGGTGC	540
Qy	714	ACCCGCTGGGGCGCGGCCCTTCGCTGGCGCGCGCTGGCCCTTGGAGACTGTGATGGCTG	773
Db	541	ACCCGCTGGGGCGCGGCCCTTCGCTGGCGCGCGCTGGCCCTTGGAGACTGTGATGGCTG	600
Qy	774	CTTGGGCTCATGGCGCGCCCTTGGCGACTGCGCCCTTGACACTGCACGCGACAGCTTCGGC	833
Db	601	CTTGGGCTCATGGCGCGCCCTTGGCGACTGCGCCCTTGACACTGCACGCGAGACTTTCGGC	660
Qy	834	TGGCGGCTCCGATCGCTGCTCTGCCATGACGCGCTGCCCTTGAGAGCACAGGCTCTCC	893
Db	661	TGGCGGCTCCGATCGCTGCTCTGCCATGACGCGCTGCCCTTGAGAGCACAGGCTCTCC	720
Qy	894	ACTGGCAACGGCGCTTCACCTGCCTGGGCGCTGTGGGCTGTTTCCCTGCCCTCGTGGCCA	953
Db	721	ACTGGCAACGGCGCTTCACCTGCCTGGGCGCTGTGGGCTGTTTCCCTGCCCTCGTGGCCA	780
Qy	954	TGCTGTGCTGTACGGGGCACCCCTTGACACTGGCGGCGCGCGCGCGGCTACGGCC	1013
Db	781	TGCTGTGCTGTACGGGGCACCCCTTGACACTGGCGGCGCGCGCGGCTACGGCC	840
Qy	1014	ACGCGTGTAGGCTGACCGCAGTGTGCTGGCCTCCGCGCTGGCCTTCTGCTGCCACGA	1073
Db	841	ACGCGTGTAGGCTGACCGCAGTGTGCTGGCCTCCGCGCTGGCCTTCTGCTGCCACGA	900
Qy	1074	ACCTGCTGTGCTGTGCTATTACTCGGACCGAGCGCCAGCGCCTGGGGCAACCTCTATG	1133
Db	901	ACCTGCTGTGCTGTGCTATTACTCGGACCGAGCGCCAGCGCCTGGGGCAACCTCTATG	960
Qy	1134	GTGCTACGTGCCACGCTGGCGCTGAGCACCTCAACAGCTGGTGGATCCCTTCATCT	1193
Db	961	GTGCTACGTGGCCAGCCTTGGCGCTGAGCACCTCAACAGCTGGTGGATCCCTTCATCT	1020
Qy	1194	ACTACTACGTGTGGCGAGTTTCAGGGACAAGTGGGGCAGGCTCTTCCAACGGCTCGC	1253
Db	1021	ACTACTACGTGTGGCGAGTTTCAGGGACAAGTGGGGCAGGCTCTTCCAACGGCTCGC	1080
Qy	1254	CGGGGACACCGTGGCCTCCAAGGCTCTGCGGAAGGGGGCAGCCGGGGCATGGGCACC	1313
Db	1081	CGGGGACACCGTGGCCTCCAAGGCTCTGCGGAAGGGGGCAGCCGGGGCATGGGCACC	1140
Qy	1314	ACTCCTTTGCTCCAGTGCACAAAGTGGGAAGGCTGTACTGGGTGCAACAGGGTCCC	1373
Db	1141	ACTCCTTTGCTCCAGTGCACAAAGTGGGAAGGCTGTACTGGGTGCAACAGGGTCCC	1200
Qy	1374	TTCCCCACTTCACGTCTCTCTGGGACCTCAGAAATGTACCTATTATTGGAAATAGGGTT	1433
Db	1201	TTCCCCACTTCACGTCTCTCTGGGACCTCAGAAATGTACCTATTATTGGAAATAGGGTT	1260
Qy	1434	GTTACAACTGTACTAGCGGAGTCACTTTGGGAAGGTTGGGCTTTACATCCAGTGTGG	1493
Db	1261	GTTACAACTGTACTAGCAGAGTCACTTTGGGAAGGTTGGGCTTTACATCCAGTGTGG	1320
Qy	1494	GTGGTGTCTCATAAATTAAGNAGGCGCAGCCCTGGTGGCTCAGGCTGTAAATCCAGC	1553
Db	1321	GTGGTGTCTCATAAATTAAGNAGGCGCAGCCCTGGTGGCTCAGGCTGTAAATCCAGC	1380
Qy	1554	ACTTTAAGAGGCCAAGCGGATGGATCACTTCGACGCCAGGAGTTCAACACCGAGCTGAGC	1613
Db	1381	ACTTTAAGAGGCCAAGCGGATGGATCACTTCGACGCCAGGAGTTCAACACCGAGCTGAGC	1440
Qy	1614	AACATGGTAAACCCCATCTCTACCAAAAATACAAAAATAGCTGGGCTTGGTGGCTGGC	1673

Db 1441 AACATGGTAAACCCCATCTCTACCAAAATACAAAATTAAGTGGGCTTGCTGGCTGGC 1500
 Qy 1674 GCCTGTATATCCAGCTACTACTCAGGAGCTAGGCA 1707
 Db 1501 GCCTGTATATCCAGCTACTCANGAGCTAGGCA 1534

RESULT 3
 AF080215 1361 bp mRNA ROD 06-AUG-1998
 LOCUS Mus musculus protease-activated receptor 4 (Par4) mRNA, complete cds.
 ACCESSION AF080215
 NID 93396082
 VERSION AF080215.1 GI:3396082
 KEYWORDS house mouse.
 SOURCE Mus musculus
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 1361)
 AUTHORS Kahn, M.L., Hammes, S.R., Botka, C. and Coughlin, S.R.
 TITLE Gene and locus structure and chromosomal localization of the protease-activated receptor gene family
 J. Biol. Chem. (1998) In press
 REFERENCE 2 (bases 1 to 1361)
 AUTHORS Kahn, M.L., Zheng, Y.W., Huang, W., Bigornia, V., Zeng, D., Moff, S., Farese, R.V., Tan, C. and Coughlin, S.R.
 TITLE A dual thrombin receptor system for platelet activation
 Nature (1998) In press
 REFERENCE 3 (bases 1 to 1361)
 AUTHORS Kahn, M.L. and Coughlin, S.R.
 TITLE Direct Submission
 JOURNAL Submitted (24-JUL-1998) CVRI, UCSF, 513 Parnassus, San Francisco, CA 94143-1030, USA

FEATURES
 source Location/Qualifiers
 1..1361
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /chromosome="8"
 /map="8B3.3"
 1..1361
 /gene="Par4"
 1..1191
 /gene="Par4"
 /note="G protein-coupled receptor; thrombin receptor"
 /codon_start=1
 /product="protease-activated receptor 4"
 /protein_id="AAC28861.1"
 /db_xref="PID:g3396083"
 /db_xref="GI:3396083"
 /translation="MCNPILYPLVLGISLAEGIQPSIYDDVSTRSGHEGPIGPTV
 VELKPKSSKPNRPGPGRFCANDSTLELPASSQALLGNWSTRLVPALYGLVYAA
 GLPANGALVMTATVRLPRLPSTILLTNLAVDSLLALVPPRLALHURQORWPEGA
 CRVATAYLYGATVSVLLAAVSDRYLALVHLRLARALRGQRLLTGCLCVAMLSAAT
 LALPLTHRQTFRAGSDRMCLHDALPLTEQTSWRPAFICLAVLGCFVPLLAGLCL
 GATURLAANGQRYSHALRLTALVLFSAVASETPSNVLLVLYHNSPPEAMGNLYGAW
 VPSLALSTLNSCDVPFIYVVSHEFRKVRAMLCRQPEASSSSQASREAGSRGTAICS
 STILL"

BASE COUNT 234 a 416 c 394 g 317 t
 ORIGIN

Query Match 13.9%; Score 681.6; DB 12; Length 1361;
 Best Local Similarity 72.0%; Pred. No. 5.6e-103;
 Matches 949; Conservative 0; Mismatches 329; Indels 40; Gaps 3;

Qy 179 TGGGGGCGACTCTCCTCTGGCCCTGGTGGGGTTCAGCCTCTCTGGGGGCGACCCAG 238
 Db 7 TGGCGCTGCTGCTATCCCTTTGGTGTGGGGCTCACATCAGCCTGGCAGGGCATCCAG 66

Qy 239 ACCCCACGCGTCTACGACGAGAGCGGGGAGCACCCGCGAGGTGTGTATGACGACACGC----- 294

JOURNAL	Submitted (10-DEC-1996) J.E. Lamerdin, Human Genome Center, Lawrence Livermore National Laboratory, 7000 East Ave, Livermore, CA, USA, 94551 janeacgt.llnl.gov ow@ornak.llnl.gov
COMMENT	GSDB:S:1010600. Human Genome Center Biology and Biotechnology Research Program Lawrence Livermore National Laboratory 7000 East Avenue Livermore, CA 94550 USA constructed at LLNL from flow-sorted chromosomes from hybrid 5HL2-B, which carries chromosome 19 as its only human chromosome.
FEATURES	Location/Qualifiers
source	1. .41936 /organism="Homo sapiens" /db_xref="taxon:9606" /chromosome="19" /cell_line="5HL2-B" /cell_type="fibroblast" /clone="R30879" /clone_lib="LL19NC03 R chromosome 19-specific cosmid library"
repeat_region	complement(1. .34) /note="repeat match = HSAL15306; putative"
repeat_region	/rpt_family="Alu" complement(811. .1099) /note="repeat match = HSAL00208; putative"
repeat_region	/rpt_family="Alu" 1335. .1479 /note="repeat match = HSAL05512; putative"
repeat_region	/rpt_family="Alu" 1367. .2305 /note="repeat match = HSAL02043; putative"
repeat_region	/rpt_family="Alu" complement(3093. .3356) /note="repeat match = HSAL02254; putative"
repeat_region	/rpt_family="Alu" 4030. .4168 /note="repeat match = HSAL03448; putative"
repeat_region	/rpt_family="Alu" 4031. .4288 /note="repeat match = HSAL01019; putative"
repeat_region	/rpt_family="Alu" 4435. .4533 /note="repeat match = HSAL02522; putative"
repeat_region	/rpt_family="Alu" 4449. .4565 /note="repeat match = HSAL01715; putative"
repeat_region	/rpt_family="Alu" 4926. .5218 /note="repeat match = HSAL06523; putative"
repeat_region	/rpt_family="Alu" complement(5804. .6091) /note="repeat match = HSAL04878; putative"
repeat_region	/rpt_family="Alu" complement(6051. .6102) /note="repeat match = HSAL04290; putative"
repeat_region	/rpt_family="Alu" 7131. .7434 /note="repeat match = HSAL02525; putative"
repeat_region	/rpt_family="Alu" 7137. .7589 /note="repeat match = HSAL00829; putative"
gene	/rpt_family="Alu" 7911. .16824 /gene="LISCH7"
CDS	join(<7911. .8030,10891. .10947,11512. .11658,15325. .15498,15802. .15933,16075. .16700,16790. .16824) /gene="LISCH7" /function="human homolog of Mus musculus B6CBA Lisch7 mRNA"
	/note="putative" /codon_start=2

85040 in this sequence. This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.

This sequence was generated from part of bacterial clone contigs of human chromosome 22, constructed by the Sanger Centre Chromosome 22 Mapping Group. Further information can be found at

<http://www.sanger.ac.uk/HGP/Chr22>

78F24 is from the library RPC11 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong. For further details see <http://bacpac.med.buffalo.edu/VECTOR:pcYPAC2>.

FEATURES

```
source
1..145414
  /organism="Homo sapiens"
  /db_xref="taxon:9606"
  /chromosome="22"
  /clone="78F24"
  /map="q12.1-12.3"
  /clone_lib="RPC11"
repeat_region
1..835
  /note="L1 repeat: matches 3208..4060 of consensus"
  1219..1514
  /note="AluX repeat: matches 1..302 of consensus"
  1575..1770
  /note="L1 repeat: matches 448..4647 of consensus"
  1775..2076
  /note="AluX repeat: matches 1..302 of consensus"
  2100..2247
  /note="L1 repeat: matches 4688..4838 of consensus"
  2108..2110
  /note="clone 934P1; acc in this entry; substitution"
  /replace="agg"
  complement(2251..2553)
  /note="AluX repeat: matches 299..1 of consensus"
  2390..2392
  /note="clone 934P1; ata in this entry; substitution"
  /replace="aca"
  2936..3262
  /note="L1ME1 repeat: matches 1..336 of consensus"
  complement(3286..3588)
  /note="AluJ repeat: matches 302..2 of consensus"
  3589..4123
  /note="L1ME2 repeat: matches 328..892 of consensus"
  4354..4569
  /note="L1MD2 repeat: matches 453..670 of consensus"
  4354..4439
  /note="L1P47 repeat: matches 453..557 of consensus"
  4618..4981
  /note="L1 repeat: matches 3159..3546 of consensus"
  4995..5292
  /note="AluSg repeat: matches 1..299 of consensus"
  5298..5596
  /note="AluX repeat: matches 3..300 of consensus"
  5638..5844
  /note="L1 repeat: matches 3650..3858 of consensus"
  5804..6123
  /note="L1MC3 repeat: matches 35..312 of consensus"
  6124..6431
  /note="AluX repeat: matches 1..301 of consensus"
  6485..6628
  /note="L1ME2 repeat: matches 364..518 of consensus"
  complement(6676..6972)
  /note="AluSg repeat: matches 300..1 of consensus"
  6994..7127
  /note="TIGER1 repeat: matches 55..186 of consensus"
  complement(7129..7424)
  /note="AluSg repeat: matches 303..2 of consensus"
  7426..8399
  /note="TIGER1 repeat: matches 178..1183 of consensus"
  8400..8700
  /note="AluJ repeat: matches 1..300 of consensus"
```

```
repeat_region
  complement(8701..8991)
  /note="AluJ repeat: matches 288..2 of consensus"
  8993..9139
  /note="TIGER1 repeat: matches 1168..1306 of consensus"
  9268..9489
  /note="TIGER1 repeat: matches 1545..1766 of consensus"
  9491..9800
  /note="AluX repeat: matches 1..302 of consensus"
  9808..10255
  /note="TIGER1 repeat: matches 1757..2224 of consensus"
  complement(10258..10393)
  /note="AluJ repeat: matches 301..190 of consensus"
  10399..10622
  /note="AluJ repeat: matches 72..294 of consensus"
  complement(10623..10670)
  /note="AluJ repeat: matches 288..241 of consensus"
  complement(10676..10855)
  /note="AluSg repeat: matches 177..1 of consensus"
  10858..11035
  /note="TIGER1 repeat: matches 2217..2418 of consensus"
  11034..11373
  /note="L1MB8 repeat: matches 529..874 of consensus"
  complement(11390..11691)
  /note="AluX repeat: matches 302..1 of consensus"
  11696..11749
  /note="L1MB8 repeat: matches 863..920 of consensus"
  complement(11857..12037)
  /note="MIR repeat: matches 259..47 of consensus"
  12094..12204
  /note="MER5A repeat: matches 2..118 of consensus"
  complement(12202..12239)
  /note="L1 repeat: matches 4967..4931 of consensus"
  12240..12267
  /note="14 copies 2 mer tt 89% conserved"
  complement(12275..12323)
  /note="MIR2 repeat: matches 141..93 of consensus"
  complement(12361..12491)
  /note="L1P2 repeat: matches 893..763 of consensus"
  12519..12974
  /note="L1 repeat: matches 4931..5390 of consensus"
  12832..13729
  /note="L1P3 repeat: matches 1..893 of consensus"
  13707..13803
  complement(14081..141115)
  /note="AluJ repeat: matches 302..1 of consensus"
  14461..14631
  /note="MIR repeat: matches 73..248 of consensus"
  14912..14914
  /note="clone 934P1; gcg in this entry; substitution"
  /replace="ggg"
  complement(15293..15594)
  /note="AluX repeat: matches 302..12 of consensus"
  complement(15636..15708)
  /note="MIR2 repeat: matches 146..74 of consensus"
  16169..16304
  /note="MIR repeat: matches 13..159 of consensus"
  16486..16488
  /note="clone 934P1; gag in this entry; substitution"
  /replace="ggg"
  16557..16855
  /note="AluJ repeat: matches 1..301 of consensus"
  17057..17473
  /note="L1 repeat: matches 4756..5191 of consensus"
  17483..17760
  /note="AluX repeat: matches 24..302 of consensus"
  17773..17970
  /note="L1 repeat: matches 5184..5390 of consensus"
  17830..18397
  /note="L1MB6 repeat: matches 1..574 of consensus"
  18167..18169
  /note="clone 934P1; tct in this entry; substitution"
  /replace="tat"
  18398..18517
  repeat_region
```



```
USA
COMMENT      Sequence submitted by:
DOE Joint Genome Institute.
FEATURES
  source      Location/Qualifiers
              1. .134506
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /map="5p"
                /chromosome="5"
  source      join(1. .29755,78169. .82395,123646. .134506)
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /chromosome="5"
                /map="5p"
                /clone="BAC 7g12"
                /note="LBNL H126"
                complement(748. .891)
  misc_feature /note="GRAIL 2 excellent exon, frame 2"
  misc_feature 2635. .2694
  repeat_region /note="GRAIL 2 excellent exon, frame 2"
  repeat_region /rpt_family="MLT1"
  repeat_region complement(2883. .2994)
  repeat_region /rpt_family="Alu"
  repeat_region /rpt_family="Alu"
  repeat_region complement(5036. .5148)
  repeat_region /rpt_family="MIR"
  repeat_region 9194. .9650
  repeat_region /rpt_family="MER42"
  repeat_region 9745. .10340
  repeat_region /rpt_family="MER42"
  repeat_region 10346. .10624
  repeat_region /rpt_family="Alu"
  repeat_region 10674. .10876
  repeat_region /rpt_family="MER42"
  repeat_region 12173. .12472
  repeat_region /rpt_family="Alu"
  repeat_region 14779. .14955
  repeat_region /rpt_family="L1"
  repeat_region 15163. .15487
  repeat_region /rpt_family="Alu"
  repeat_region complement(19591. .19738)
  repeat_region /rpt_family="MIR"
  repeat_region 20251. .20549
  repeat_region /rpt_family="Alu"
  misc_feature complement(21622. .21774)
  repeat_region /note="GRAIL 2 excellent exon, frame 1"
  repeat_region complement(22628. .22797)
  repeat_region /rpt_family="MIR"
  repeat_region complement(23374. .23671)
  repeat_region /rpt_family="Alu"
  repeat_region 24507. .24641
  repeat_region /rpt_family="MER7"
  source      28893. .60995
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /chromosome="5"
                /map="5p"
                /clone="P1 210h6"
                /note="LBNL H52"
  repeat_region complement(34480. .34672)
  repeat_region /rpt_family="MLT1"
  misc_feature complement(34676. .34821)
  repeat_region /note="GRAIL 2 excellent exon, frame 0"
  repeat_region complement(34837. .35181)
  repeat_region /rpt_family="THE1"
  misc_feature 38746. .38854
  repeat_region /note="GRAIL 2 excellent exon, frame 1"
  repeat_region 39274. .39561
  repeat_region /rpt_family="Alu"
  repeat_region 39856. .40341
  repeat_region /rpt_family="MER1"
  repeat_region complement(40375. .40451)
  repeat_region /rpt_family="MLT1"

repeat_region 44067. .44382
/rpt_family="Alu"
repeat_region 44592. .44900
/rpt_family="Alu"
repeat_region 45213. .45338
/rpt_family="L1"
repeat_region 45576. .45708
/rpt_family="Alu"
repeat_region 46954. .47235
/rpt_family="Alu"
repeat_region 50715. .51019
/rpt_family="Alu"
repeat_region 51522. .51600
/rpt_family="MIR"
repeat_region complement(51960. .52100)
/rpt_family="MIR"
repeat_region complement(52323. .52389)
/rpt_family="MER7"
repeat_region complement(52403. .52692)
/rpt_family="Alu"
repeat_region complement(52752. .53008)
/rpt_family="Alu"
repeat_region complement(53050. .53338)
/rpt_family="Alu"
repeat_region complement(53346. .53568)
/rpt_family="MER7"
repeat_region complement(55100. .55354)
/rpt_family="L1"
repeat_region complement(55478. .55533)
/rpt_family="Alu"
repeat_region complement(56376. .56704)
/rpt_family="L1"
repeat_region 57025. .57175
/rpt_family="MLT1"
repeat_region complement(58128. .58416)
/rpt_family="Alu"
repeat_region complement(58603. .58884)
/rpt_family="L1"
repeat_region 59416. .59747
/rpt_family="MER21"
source      join(60081. .78733,82162. .124917)
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone="P1 380h11"
                /chromosome="5"
                /map="5p"
                /note="LBNL H122"
  repeat_region complement(60553. .60710)
/rpt_family="Alu"
  repeat_region complement(60718. .60914)
/rpt_family="THE1"
  repeat_region complement(61908. .62249)
/rpt_family="THE1"
  repeat_region complement(62279. .63781)
/rpt_family="MSTAR"
  repeat_region complement(63786. .64160)
/rpt_family="THE1"
  repeat_region 65491. .65778
/rpt_family="Alu"
  misc_feature complement(66977. .67113)
/rpt_family="THE1"
  repeat_region /note="GRAIL 2 excellent exon, frame 2"
  repeat_region 67119. .67157
/rpt_family="THE1"
  misc_feature complement(68013. .68120)
/rpt_family="THE1"
  repeat_region /note="GRAIL 2 excellent exon, frame 1"
  repeat_region 68346. .68651
/rpt_family="MIR"
  repeat_region 68894. .69226
/rpt_family="THE1"
  repeat_region complement(69684. .70023)
/rpt_family="THE1"
  repeat_region 70567. .70681
/rpt_family="Alu"
```

```
repeat_region complement(72274..72537)
misc_feature /rpt_family="Alu"
misc_feature /note="GRAIL 2 excellent exon, frame 2"
repeat_region complement(77047..77436)
misc_feature /rpt_family="MLN1"
misc_feature 79647..79829
misc_feature /note="GRAIL 2 excellent exon, frame 2"
misc_feature complement(82876..82937)
misc_feature /note="GRAIL 2 excellent exon, frame 0"
repeat_region 83530..83674
repeat_region /rpt_family="Alu"

Query Match 10.08; Score 489.2; DB 11; Length 134506;
Best Local Similarity 74.78; Pred. No. 1e-71;
Matches 728; Conservative 0; Mismatches 218; Indels 28; Gaps 8;

Qy 1518 AGGCGAGCGCTGGTGGCTCAGCCCTGTAATCCAGACACTTTAAAGAGGCCAAGCGGGATGG 1577
Db 53347 AAGCCCGTGTGGTGGCTCAGCTCTGTAATCCAGACACTTTGGAGGCTGAGCGGGTGG 53288
Qy 1578 ATCACTTGAGCCCGAGGAGTTCAACACCCAGCCCTGAGCAACATGTTAAACCCCATCTCTAC 1637
Db 53287 ATCACTTGAGCTCAGGAGTTCTAGACCA-CCTGGCCACATGTTGAAACCCCATCTCTGAC 53229
Qy 1638 CAATAATACAAAATATAGCTGGCTTGGTGGCTGGCGCTGTAATCCAGCTACTCAGGA 1697
Db 53228 TGAATAATACAAAATATAGCTGGCTTGGTGGCTGGCTGTAATCCAGCTACTC--GG 53171
Qy 1698 GACTGAGCAGCAAGGATCGCTTGAACCTGGAGGAGGAGTGGAGTGGAGCGGAGATTGC 1757
Db 53170 GCCTGAGCAGGAGATGTTTGAACCGGAGGAGGAGTGGAGTGGAGCGGAGATTGC 53111
Qy 1758 GCCACTGACTCCAGCCTCGCTGAGCAGAGCCCTGTCTCTAAATTAATTAATTAAT 1817
Db 53110 ACCACTGACTCCAGCCTGGGCAACAGAGCAAGACTCCATCTCAAAAAAATAAAAAA 53051
Qy 1818 TTAATCAATTTTAAAGAGCAAAAGTGACGGCCAGTGGAGTGGCTGAGCGCTATAT 1877
Db 53050 AGAAATATCAGTAAGATACAGTATATATAGCCAGGCACTGTGGCTCATGACAGTAA 52991
Qy 1878 CTCAGCACTCGGAGGCAAGATGGAGATGCTTGAAGCAGGAGTGGAGTGGAGCAGCC 1937
Db 52990 CACAGCACTTTGGAGGCAAGGAGGAG-CGATACAGGAGTGGAGTGGAGCAGCAGCC 52932
Qy 1938 TGGGCAACATAGGGGATGCCATCTCTACACACAAAAAATTTTAAATGAACAGGAT 1997
Db 52931 TGGCAACACAGTGAACCCCGTCTCTCTAAAAATACAA---AAATCAGCCAGCGGG 52876
Qy 1998 TGTGGCATGCCCTATAGTCCAGCCACTCAAGAGGCAAGGAGGAGGATCACTTGAGC 2057
Db 52875 GGTGGCAGCGCCCTGTAGTCCAGCTACTTGGGAGCTGGAGCAAGAAATGCTTGAAC 52816
Qy 2058 CTGGGAGTGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 2117
Db 52815 CCAGGAGACAGGTTGAGTGGGCAAAATCATGCCACTGCATCCCAACTGGGCAACA 52756
Qy 2118 GAGCAAGACCTTGTCTCAAAAATTAACAACTAAATTAATAAAGAGAGAGAGATAG 2177
Db 52755 GAGCAAGACTGTCTCAGAAAGAAA---GAAAGAAAAGAAAGATATAGTGGC 52700
Qy 2178 TGGTGTGGTGGTGCACACTGCAATCCAGCACTTTGGAGGCGGAGGTGGGAGATCA 2237
Db 52699 TGGGCGTGGTGGCTCACACTTGTATCCAGCACTTTGGAGGCGGAGAGTGGTGGATCA 52640
Qy 2238 TCTGAGGCGAGGAGTTCAAGACCGCTGGCTACATGCTGAATCTCTCTACCAA 2297
Db 52639 --CAAGGTGAGAGTTCAAGACCGCTGGCGGAGTGGTGGGAGGAGTGGTGGTGGTGG 52582
Qy 2298 AATAC-AAAAATAGCCAGGCGTGGTGGTGGGCACTTCTACTGAGGAGTGGCCACCCAG 2356
Db 52581 AATACAAAATATAGCCAGGCGTGGTGGGCGGCTGTAGT-----CCGAG 52535
```

```
Qy 2357 CTACTGGGGAGGCTGAGTCAAGGAGTCACTGAACTGGAGCGGAGGCTGGGTGAG 2416
Db 52534 CTACTCGGGAGGCTGAGGAGGAGTGGCGTGAACCTGGAGCGGAGCTTGCAGTGAG 52475
Qy 2417 CTGAGATGGTGGCCACTGCATCTCCAGCCTGGGCGAAGAGGAGCTCTCTCTCAAAAAA 2476
Db 52474 CCGAGATTTTGTCACTGCATCTCCAGCCTGGGCAACAGGAGAGCTCCGCTCTCAAAAAA 52415
Qy 2477 GAGAAGAGGAGGAGG 2490
Db 52414 AAAAAAGAAAAAG 52401

RESULT 9
AC005551 LOCUS 42416 bp DNA PRI 30-DEC-1998
DEFINITION Homo sapiens chromosome 19, cosmid R26529, complete sequence.
AC005551
NID G3482904
VERSION AC005551.1 GI:3482904
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 42416)
AUTHORS Lamerdin,J.E., McCready,P.M., Skowronski,E., Adamson,A.W.,
Burkhardt-Schultz,K., Gordon,L., Kyle,A., Ramirez,M., Stilwagen,S.,
Phan,H., Velasco,N., Do,L., Regala,W., Terry,A., Danganan,L.,
Poundstone,P., Christensen,M., Georgescu,A., Avila,J., Liu,S.,
Attix,C., Andrise,T., Frankheim,M., Amico-Keller,G., Coefield,J.,
Darte,S., Lucas,S., Bruce,R., Thomas,P., Quan,G., Krommiller,B.,
Arellano,A., Ow,D., Nolan,M., Trong,S., Kobayashi,A., Olsen,A.S.
and Carrano,A.V.
TITLE Sequence analysis of a 3.5 Mb contig in human 19p13.3 containing a
serine protease gene cluster
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 42416)
AUTHORS Lamerdin,J.E.
TITLE Direct Submission
JOURNAL Submitted (28-AUG-1998) Joint Genome Institute, Lawrence Livermore
National Laboratory, 7000 East Ave., Livermore, CA 94551, USA
COMMENT Map and sequence oriented from p telomere to centromere. Cosmid
R26529 is separated from cosmid F22676 to the left by approximately
1.5 kb, and is separated from cosmid R3374 to the right by
approximately 1.5 kb. Additional chromosome 19 map and sequence
information may be obtained at:
http://www-bio.llnl.gov/bbrp/genome/genome.html.
FEATURES
source
1..42416
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="R26529"
/chromosome="19"
/map="19p13.3 between CDC34 and D19S342"
/cell_line="5HL2-B"
/clone_lib="LL19NC03 R chromosome 19-specific cosmid
library"
/note="Cosmid library constructed at LLNL from flow-sorted
chromosomes from human-hamster hybrid 5HL2-B, which
carries chromosome 19 as its only human chromosome."
repeat_region 56..162
/rpt_family="MIR"
repeat_region 187..473
/rpt_family="AluSx"
repeat_region 476..610
/rpt_family="AluJo"
mrna <834..11329
/gene="NFIC"
/product="H.sapiens mRNA for NF1/CAAT-binding
transcription factor 5 (CTF5)"
<834..11303
/gene="NFIC"
```



```

/product="Human mRNA for CCAAT-box binding transcription
factor CTF-1 (syn. CTF/NFI or CTF or NF-I or NF-1)"
<834. .11329
/gene="NFI"
/note="nuclear factor I/C (CCAAT-binding transcription
factor)"
join(<834. .1018,11104. .11121)
/gene="NFI"
/note="NFI /CAAT-binding transcription factor 5"
/codon_start=3
/product="CTF5"
/protein_id="AAC33192.1"
/db_xref="PID:G3482907"
/db_xref="GI:3482907"
/translation="TARSPHSSALHFTTSTLPQTASTYFPHTAIRYPHLPQDPL
KDLVSLACDPASQPGSPKSWLG"
CDS
join(<834. .1018,2115. .2268,4902. .4987,11104. .11121)
/gene="NFI"
/note="CAAT BOX-BINDING TRANSCRIPTION FACTOR 1"
/codon_start=3
/product="NFI1_HUMAN"
/protein_id="AAC33191.1"
/db_xref="PID:G3482906"
/db_xref="GI:3482906"
/translation="TARSPHSSALHFTTSTLPQTASTYFPHTAIRYPHLPQDPL
KDLVSLACDPASQPGSPKSWLG"
CDS
join(<834. .1018,4902. .4952)
/gene="NFI"
/note="CAAT BOX-BINDING TRANSCRIPTION FACTORS 2 AND 3"
/codon_start=3
/product="NFI2_HUMAN"
/protein_id="AAC33190.1"
/db_xref="PID:G3482905"
/db_xref="GI:3482905"
/translation="TARSPHSSALHFTTSTLPQTASTYFPHTAIRYPHLPQDPL
KDLVSLACDPASQPGSPKSWLG"
misc_feature
834. .1018
/gene="NFI"
/note="DDS similarity to N20996 yx57g02.s1 Homo sapiens
cDNA clone 265874 3' similar to gb:X12492 CCAAT BOX-
BINDING TRANSCRIPTION FACTOR 1 (HUMAN); (482. .298); 96%
identity."
/misc_feature
899. .1018
/gene="NFI"
/note="predicted exon, program: grail2exons_human.1.3,
frame: 1, quality: excellent, score: 89.000--(931. .1018)
DDS similarity to AA282279 zt12b08.r1 NCI_CGAP_GCB1 Homo
sapiens cDNA clone IMAGE:712887 5' similar to gb:X12492
CCAAT BOX-BINDING TRANSCRIPTION FACTOR 1 (HUMAN); (1. .88);
100% identity.--Additional EST matches: AI083745"
repeat_region
1306. .1504
/rpt_family="AluJo"
repeat_region
1988. .2055
/rpt_family="GC_rich"
misc_feature
2115. .2268
/gene="NFI"
/note="DDS similarity to AA282279 zt12b08.r1 NCI_CGAP_GCB1
Homo sapiens cDNA clone IMAGE:712887 5' similar to
gb:X12492 CCAAT BOX-BINDING TRANSCRIPTION FACTOR 1
(HUMAN); (89. .242); 100% identity.--Additional EST
matches: AI094574"
repeat_region
3292. .3389
/rpt_family="MIR"
repeat_region
3753. .3785
/rpt_family="LINE2"
repeat_region
3786. .3795
/rpt_family="MIR"
repeat_region
4148. .4188
/rpt_family="MIR"
misc_feature
4902. .4987
/gene="NFI"
/note="DDS similarity to overlapping ESTs:-N20996

```

```

yx57g02.s1 Homo sapiens cDNA clone 265874 3' similar to
gb:X12492 CCAAT BOX-BINDING TRANSCRIPTION FACTOR 1
(HUMAN); (297. .211); 97% identity.--AA282279 zt12b08.r1
NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:712887 5'
similar to gb:X12492 CCAAT BOX-BINDING TRANSCRIPTION
FACTOR 1 (HUMAN); (243. .328); 100% identity."
/rpt_family="MIR"
/complement(5259. .5352)
/rpt_family="MIR"
/complement(5696. .5751)
6235. .6341
/rpt_family="MIR"
/gene="NFI"
/note="DDS similarity to AA282279 zt12b08.r1 NCI_CGAP_GCB1
Homo sapiens cDNA clone IMAGE:712887 5' similar to
gb:X12492 CCAAT BOX-BINDING TRANSCRIPTION FACTOR 1
(HUMAN); (329. .435); 100% identity.--"
/complement(6401. .6765)
/gene="NFI"
/note="DDS similarity to AA282217 zt12b08.s1 NCI_CGAP_GCB1
Homo sapiens cDNA clone IMAGE:712887 3'; Score: 704
Identity: 361/363 (99%).
/complement(7093. .7146)
/rpt_family="LINE2"
8001. .8053
/rpt_family="MIR"
8679. .8774
/rpt_family="GGGA)n"
/complement(8868. .9165)
/rpt_family="AluSx"
9195. .9493
/rpt_family="AluSx"
9524. .9790
/rpt_family="AluJo"
9793. .10094
/rpt_family="AluSx"
/complement(10099. .10125)
/rpt_family="AT_rich"
10126. .10421
/rpt_family="AluSx"
10452. .10752
/rpt_family="AluSx"
10800. .10905
/rpt_family="MER53"
11104. .11314
/gene="NFI"
/note="DDS similarity to N20996 yx57g02.s1 Homo sapiens
cDNA clone 265874 3' similar to gb:X12492 CCAAT BOX-
BINDING TRANSCRIPTION FACTOR 1 (HUMAN); (210. .1); 98%
identity.--Additional EST matches: AI094574, AI083745"
11286. .11334
/rpt_family="GAA)n"
/complement(11504. .11534)
/rpt_family="GC_rich"
/complement(11586. .12033)
/note="BLASTN similarity to AI089816 (15. .462); match:
0.98, score: 3.6e-180; database searched: month.na;
qalico9.x1 NCI_CGAP_Brn23 Homo sapiens cDNA clone
IMAGE:1686448 3' similar to contains MSRI.t3 MER22
repetitive.
12659. .13782
/note="DDS similarity to overlapping ESTs:-(12659. .13136)
AA497129 aa42c05.r1 Soares NHMPu S1 Homo sapiens cDNA
clone 823592 5'; Score: 751 Identity: 436/491
(88%).--(12899. .13296) N45485 yy59c07.r1 Homo sapiens cDNA
clone 277836 5'. Score: 706 Identity: 381/402
(94%).--(13506. .13074) AA071297 zm73d09.s1 Stratagene
neurcpeithelium (#937231) Homo sapiens cDNA clone 531281
3'; Score: 709 Identity: 411/442 (92%).--(13782. .13308)
AA497062 aa42d05.s1 Soares NHMPu S1 Homo sapiens cDNA

```

Query Match 10.0%; Score 489; DB 11; Length 42416;
Best Local Similarity 74.2%; Pred. No. 1.3e-71;
Matches 722; Conservative 0; Mismatches 220; Indels 31; Gaps 7;

Qy	1519	GGCCAGCGCTGGTGGCTCAGCGCTGTAAATCCAGCAGCTTTAAGAGCGCAAGCGGATGGA	1578
	1519	GGCCAGCGCTGGTGGCTCAGCGCTGTAAATCCAGCAGCTTTAAGAGCGCAAGCGGATGGA <td>1578</td>	1578
Db	9793	GGCTGGCCACAGTGGCTCAGCGCTGTAAATCCAGCAGCTTTGGGAGCGCAAGGTGGCGGA	9852
Qy	1579	TCAGCTTGGAGCCAGGAGTTCACACACAGCCTGAGCAACATGGTAAACCCCACTCTCTACT	1638
Db	9853	TCAGCTTGGAGCCAGGAGTTCACACACAGCCTGAGCAACATGGTAAACCCCACTCTCTACT	9912
Qy	1639	AAAAATCAAAAAATTAGCTGGCTTGGTGGCTGGCGCCTGTATCCAGCTACTCAGGAG	1698
Db	9913	AAGAAATCAAAAAATTAGCTGGCTTGGTGGCGCCTGTATCCAGCTACTCAGGAG	9972
Qy	1699	ACTGAGCAGAGATCGCTTGAACCTGGGAGCGAGGTTGCAGTGAGCCAGATGGG	1758
Db	9973	GCTAAGGCACAGGATCACTTGAACCCAGGAGCGGAGTGTGTAGAGAGCCAGATGACA	10032
Qy	1759	CCAGTGGACTCAGCGCTGGTGACAGAGCGCTGTCTAAATTAATTAATTAATTAAT	1818
Db	10033	CCAGTGGACTCAGCGCTGGTGACAGAGCGCTGTCTAAATTAATTAATTAATTAAT	10092
Qy	1819	TAATTCAA---TTTTAAAAAGAGCAAGTGGTGGCGCAGTGGCTGACGCTCTATA	1875
Db	10093	GAGAAGAAATAATTTTAAATATAAATAAAGGCCAGCGAGTGGCTCAGCGCTGTA	10152
Qy	1876	ATCTCAGCACTCTGGAGGCCAGATGGAGGATGTCTTGAAGCCAGGAGTTGGGACGAG	1935
Db	10153	ATCCAGCACTTTGGAGCGCGAGCGCGGGGAT-CACGAGGTCAAGAGAGTGAGACCAT	10211
Qy	1936	CCTGGCAACATAGGGGATCCCATCTCTACACACAAAAAATTTTTTAATGAACGAGC	1995
Db	10212	CCTGGCAACATAGGGGATCCCATCTCTCTCTAC-----TAAATAATAAATAATAGCTGAGT	10266
Qy	1996	ATTGTGGCATCGCCTATAGTCCAGCACTCAAGAGGACAGCGGGAGGATCACTTGA	2055
Db	10267	GTGTGGCATCGCCTGTATCCAGCTACTCGGAGGCTGAGCGAGAGATCACTTGA	10326
Qy	2056	GCCTGGAGGTTGGTTGAGCAGTGATGTATGTACCAGCTGCATCCAGCCTGGGCAA	2115
Db	10327	ACCTGGAGGCGAGAGTTGCAATGAGCGGAGATCATGCCACTGCATCCAGCCT-GGCCAA	10385
Qy	2116	CAGACGAGACCTTGTCTCAAAATAAACAACATAAAT-----TAAAAAAGAGAGAC	2168
Db	10386	CAGACGAGACCTTGTCTCAATTAATAAATAAATAATAATAATAAATAAATAAATAA	10445
Qy	2169	GAGAGATAGTGGTGGTGGCTCAGACCTGCAATCCAGCACTTTGGAGGCGGAGTG	2228
Db	10446	GTTCAGACTGGGAGTGGCTCAGACCTGTATTTCCAGCACTTTGGGAGCGGAGGAA	10505
Qy	2229	GGCAGATCATCTGAGGCGGAGGTTCAAGACCAAGCCTGGCTAACATGGTGAATCCTATC	2288
Db	10506	GGTGGATCACTTGGTTCAGGCGTTTGAGACCAAGCCTGGGCAACATGGTGAACCCCATC	10565
Qy	2289	TCTACCAAAAATACAAAAATTAGCAGGCGTGGTGGCGCACTTACTGGGAGGTGC	2348
Db	10566	TCTACTAAAAATACAAAAATTATCCAGGTGGCGGTGGAGCGCTGTAAT-----	10616
Qy	2349	CCAGCCAGCTACTGGGAGGCTGAGTCAGAGAGATCGCTTGAACCTGGAGCGGAGGTT	2408
Db	10616	---CCAGCTACTGGGAGGCTGAGTAGGAGAGATTTCTTGAACCCAGGAGCGAGGTT	10672
Qy	2409	CGGCTAGCTGAGATGGTGGCACTGCATCCAGCTGGGCGGA-AAGAGGGAGCTCTGCTC	2467
Db	10673	GTAGTGAGCAATATACGCCAATTGCATCCAGCTGGGCAACAGAGCGAATCCATC	10732
Qy	2468	CAAAAAAAGAGA	2480
Db	10733	TCAAAAAATAAAA	10745
RESULT	10		
LOCUS	HS216E10	122320 bp	DNA
	HS216E10		PRI
			10-JUN-1999

repeat_region	/note="Alusx repeat: matches 3. .295 of consensus" 19560. .19867
repeat_region	/note="Aluy repeat: matches 1. .306 of consensus" 19871. .20176
repeat_region	/note="Alu7b8 repeat: matches 1. .313 of consensus" 20179. .20424
repeat_region	/note="Alusx repeat: matches 1. .248 of consensus" 20499. .20679
repeat_region	/note="Alusg repeat: matches 1. .175 of consensus" 20722. .20837
repeat_region	/note="L1MC5 repeat: matches 7276. .7410 of consensus" 20839. .20986
repeat_region	/note="FRAM repeat: matches 1. .151 of consensus" 21002. .21292
repeat_region	/note="Alusg repeat: matches 1. .299 of consensus" 21294. .21595
repeat_region	/note="AlusC repeat: matches 1. .301 of consensus" 21603. .21782
repeat_region	/note="Alusx repeat: matches 140. .307 of consensus" 21783. .22094
repeat_region	/note="Aluy repeat: matches 1. .309 of consensus" 22095. .22230
repeat_region	/note="Alusx repeat: matches 1. .140 of consensus" 22251. .22361
repeat_region	/note="L1MC5 repeat: matches 7401. .7517 of consensus" 22439. .22487
repeat_region	/note="AluJ0 repeat: matches 7. .57 of consensus" 22488. .22785
repeat_region	/note="Alusx repeat: matches 5. .303 of consensus" 22786. .23039
repeat_region	/note="AluJ0 repeat: matches 57. .292 of consensus" 23040. .23187
repeat_region	/note="L1ME3A repeat: matches 5967. .6110 of consensus" 23429. .23647
repeat_region	/note="AluJb repeat: matches 1. .217 of consensus" 23785. .23843
repeat_region	/note="L1M4 repeat: matches 5291. .5350 of consensus" 23844. .24147
repeat_region	/note="Alusg repeat: matches 1. .307 of consensus" 24148. .24326
repeat_region	/note="L1M4 repeat: matches 5350. .5522 of consensus" 24356. .24482
repeat_region	/note="L1MB5 repeat: matches 6050. .6175 of consensus" 24493. .24798
repeat_region	/note="L1MB5 repeat: matches 5791. .6100 of consensus" 24814. .25120
repeat_region	/note="Alu7b8 repeat: matches 1. .317 of consensus" 25139. .25436
repeat_region	/note="AlusC repeat: matches 1. .298 of consensus" 25441. .25563
repeat_region	/note="FLAMJ repeat: matches 1. .126 of consensus" 25564. .25607
repeat_region	/note="L1MB4 repeat: matches 6086. .6131 of consensus" 25608. .25647

Query Match	9.9%;	Score 486;	DB 10;	Length 122320;
Local Similarity	72.8%;	Pos. No. 3.5e-71;		
Matches	732;	Conservative	220;	Indels 54; Gaps
1519	GGCCAGGCGCTGTGGCTCAGCGCTGTATCCAGCACATTTAAGAGGCCAAAGCGGATGGA	1578		
88177	GGCAGGGCGATGTGGCTCATGTCTGTATCCAGCACATTTGGGAGGCCAAGATGGGTGGA	88118		
1579	TCACCTTAGGCCCGCAGGAGTTTCAACACAGCCGTGAGCAACATGGTAAAAACCCCTCTTACC	1638		
88117	TCACCTTAGGCTCAGGAGTTTCCAGACAGCGCTGGCCAAACATGSCAAAAACCCCTCTTACT	88058		
1639	AAAAATACAAAAATTAGCTGGCGCTTGGCTGGCGCCCTGTAAATCCACGACTACTCAGAG	1698		
88057	AAAAATACAAAAATTAGCCAGCGCATCATGGCGCATTCGCTGTAAATCCACGACTACTCAGAG	87998		
1699	ACTGAGGCAGAGGATCGGTTTGAACCTCGGAGGCAGAGGTTGCAGTCAGCCGAGATTGCG	1758		
87997	GCTGAGGCACAGAAATTGTTTGAACCCAGGAGGCGAGGTTTSCATGAGCCAAAGATCAG	87938		


```
QY 1627 CCATCTCTACCAAAATACAAAAATAGCTGGCTGTGGCTGGCTGTAAATCCCA 1686
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 94551 CCATCTCTACTAAATAC-AAATTAATCTGTGTGGTGGCGCAGCTGTAAATCCCA 94493
QY 1687 GCTACTCAGGAGTCTGAGGAGAGGATCGCTTTGAACCTGGGAGGAGGTTGCAGTGA 1746
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 94492 GCTACTCGGAGGCTGAGGAGGAGGATCGCTTTGAACCCAGGAGGTTGTTCAGTGA 94433
QY 1747 GCGAGATTCGCGCACTGGACTCCAGCTCGCTGACAGAGAGCTGTCTCTAAATTAAT 1806
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 94432 GCGAATACGCGCACTGCTGCTCCAGCTCGCTGGTGGAGATTTGAGATCTC-----CATCT 94380
QY 1807 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1866
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 94379 CAAAGTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 94320
QY 1867 AGCCCTATATCTCAGCACTCTGGGAGGC-CAAGATGGAGGATTTGCTTGAAGCCAGGAT 1925
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 94319 AGCCCTGTAAATCCAGCACTTTGGGAGGCTGAAGTATGTGGATCACTGTGGCCAGGAT 94260
QY 1926 TTGGGAGGCTGGGCAACATAGGGGGATCCCATCTCTACACACAAAAAATTTTAA 1985
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 94259 TCAAGACCAAGCTGGGCAACATAGTGAACCCCTGTCTCTACTAAAAATAAAA-----AAA 94205
QY 1986 TGAACAGGCTGTGGCATGCGCCCTATAGTCCAGCCACTCAAGAGGCACAGCGGGAG 2045
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 94204 AATCCGGGCTGTGGTGGGCACTTTATCCAGCTACTCGGAGCTGAAGCAAGAG 94145
QY 2046 GATCACTTGAGCCTGGGAGGTTGTGTGCAGTGAGCTATGATTTACCACCTGCACCTCA 2105
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 94144 AATTACTTGAACCTGGGAGGAGGTTGCAGTAAGCCAGATCTTGTCTGCCTCACTCCA 94085
QY 2106 GCTGGGCAACAGCAAGCACTTGTCTCAAAAAATAACAACTAAATTAATAAAGAA 2165
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 94084 GCTGGGCAACAGCAAGCACTTGTCTCAAAAAATAAATAAATAAATAAATGTTGTG 94025
QY 2166 GACGAGAGATAGTGGTGTGGTGTGGCTCACACCTGCAATCCAGCACTTTGGAAGCGGAG 2225
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 94024 GGCCA-----GGCATGGTGGCTTACACCTGTAAATCCAGCACTTTGGGAGCGCAG 93974
QY 2226 GTGGGAGATCATCTGAGGCGAGGATTTCAAGACCAAGCTGGCTTAACATGGTGAATCTCT 2285
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 93973 GCAGTGGATCATCTGAGTTCAGGAGTTCTAGACCAAGCCCGGCAATACGGTGAACCC 93914
QY 2286 ATCTCTACCAAAATACAAAAATAGCCAGGCTGTGTGGGCACTGTACTGGGGAGG 2345
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 93913 ATTCTTAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 93860
QY 2346 TGCCCAACCCAGCTACTGGGAGGCTGAGTCAGGAGAAATCGCTTGAACCTGGGAGCGGAG 2405
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 93860 -----CCCATCTGCTCGGAGAGCTGAGGATGAATAATCGCTTGAATTTGGGGTGGAG 93807
QY 2406 GTTGGGCTGAGTGAATGGTGCATGCACTCCAGCTCGGCGGAAAGACGCACTCTGTCT 2465
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 93806 GTTGAGTGAATGAGTGTGCATGCTGCTCCAGCTGCTCCAGCCAGGCGGAGAGTGAAGTCTG 93747
QY 2466 TCCAAAAAAGAGAGGAGGAGGACACAGAGACACAGAGAA 2513
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 93746 TCTCAAAAAAAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 93699
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

```
RESULT 12
LOCUS HS50H1/c
DEFINITION Homo sapiens chromosome 20 clone 550H1, WORKING DRAFT SEQUENCE, in
unordered pieces.
ACCESSION AL035420
NID 95050978
VERSION AL035420.9 GI:5050978
KEYWORDS HTG; HTGS_PHASE1.
SOURCE human.
ORGANISM Homo sapiens
```

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 109613)
Skuce,C.
Direct Submission
Submitted (10-JUN-1999) Wellcome Trust Genome Campus, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humqueresanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Jun 11, 1999 this sequence version replaced gi:4741490.
IMPORTANT: This sequence is unfinished and does not necessarily
represent the correct sequence. Work on the sequence is in progress
and the release of this data is based on the understanding that the
sequence may change as work continues. The sequence may be
contaminated with foreign sequence from E. coli, yeast, vector,
phage etc. Order of segments is not known; 800 n's separate
segments. Unfinished: dj550H1 Contig_ID: 01651 acc=AL035420
Length: 74457 bp Unfinished: dj550H1 Contig_ID: 01651 acc=AL035420
acc=AL035420 Length: 34356 bp.
* NOTE: This is a 'working draft' sequence.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

FEATURES

Location/Qualifiers
1..109613
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="20"
/clone="550H1"

BASE COUNT 27126 a 25598 c 26628 g 29461 t 800 others

Query Match

Best Local Similarity 9.9%; Score 484; DB 34; Length 109613;
Matches 724; Conservative 0; Mismatches 220; Indels 28; Gaps 8;

```
QY 1519 GCGCAGGCTGTGGCTCACGCTGTAAATCCAGCACTTTAAGAGCCCAAGCGGATGGA 1578
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 95217 GCGCAGTGGTGTACTCATCTGTCTACTCCAGCACTTTGGGAGCGGCGAGTGA 95158
QY 1579 TCATTTGAGCCCGAGGTTCAACACAGCCTGAGCAACATGGTAAACCCCATCTCTACC 1638
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 95157 TCACCTGAGTTCAGGAGTTTGAACACAGCCTGGCCACACAGTGAATCTCATCTACT 95098
QY 1639 AAAATACAAAATAGCTGGCTGTGGTGGCTGGCTGTAAATCCAGCACTCTACAGAG 1698
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 95097 AAAATACAAAATAGCTGGGATGGTGGGACACAGTTTGTAGTCCAGCACTTTGGGAG 95038
QY 1699 ACTGAGGAGAGGATCGCTTGAACCTGGGAGGAGAGGTTGCAGTGCAGGAGATTGCG 1758
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 95037 GCTGAAGCAGGAGATCATTTGAACCTCAGGAGTTGGAGTTGCAGTGGCGGAGATGGA 94978
QY 1759 CCATCTGAGTCTCAGCTCGGTGACAGAGAGCTGTCTCTAAATTAATTAATTAATTAAT 1818
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 94977 CCATCTGAGTCTCAGCTCGGTGACAGAGAGCCTGTCTCAAAACATAAAATAAATAA 94918
QY 1819 TAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1878
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 94917 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 94860
QY 1879 TCAGCACTCTGGGAGGCGCAAGAT-GGAGGATTTGCTTGAAGCAGGAGTTTGGGAGCAGCC 1937
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 94859 CCAGCACTTTGGGAGGCGAGGTTGGTGGATCACCTGAGTGCACAGTTCGAGACCGCC 94800
QY 1938 TGGGCAACATAGGGGATCCCATCTCTACACACAAAAAATTTTTTAATGAACAGGAT 1997
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 94799 TGACCAACAGGGTAAACCCCGCTCTCTACTAAAAATACAA-----AAAATAGCCCTGGCT 94744
QY 1998 TGTGCGCATGCGCTATAGTCCAGCACTCAAGAGGCGACAGCGGAGGATCACTTGAGC 2057
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 94743 GTTGGGAGTGCCTGTAAATCCAGCTACTTGGGAGGCTGAGGAGGAGGATCACTTGAGC 94684
QY 2058 CTGGGAGGTTGTGTGTCAGTGAATGATGTACCAGCTGCAGCTCCAGCCTGGGCAACA 2117
```


[illegible]

Qy	1799	AATTAAATTAATTAATTAATTAATTCAC-----ATTTTAAAAAGACGAAA	1843
Db	64901	AAAAAAAAAAAAAAAAAAGAGAGAGGTGGAGAACACAGACAATACTCGGAGAAAGAAATAGCCA	64960
Qy	1844	GTGACGCCAGGTGCAGTGGCTCAGCCCTATAACTCTACGCACTCTGGGAGGCCAGATGG	1903
Db	64961	CAGAGCACAGGTGCAGTGGCTCACACCTGTAATCCCAGCACTCTGGCAGGCCGAGGTGG	65020
Qy	1904	AGGATTGCTTGAAGCCAGGAGTTTGGGACCAGCCCTGGCAACATAGGGGGATCCCATCTC	1963
Db	65021	-GCAGATCACGAGGTTCAGAGCTTCGAGCACAGCCCTGGCTAACATGCTGAACCCCTGTCTG	65079
Qy	1964	TACACACAAAAAATTTTTTAATGAACAGAGCATTTGTGGCATGCGCCCTATAGTCCCGCC	2023
Db	65080	TAC-----TAAAAATACAAAAATTAGCTGGACGTGGTGTTGTGGCTGTAGTTCCAGCT	65134
Qy	2024	ACTCAAGAGGCACAGCGGGAGGATCACTTCAGCCTCGGAGGTTCTGGTTCGAGTGAGCT	2083
Db	65135	ACTCAGGAGGCTGAGGCACAGAATCACTTGACCCGGAGGTTGGAGGTTGCAGTGAGCC	65194
Qy	2084	ATGATTGTACCATGCACTCCAGCCTCGGCAACAGAGCAAGCACTTGTCTCAAATAATAA	2143
Db	65195	AAGATCGGCCATTTGTAATCTAGCTGGGCAATAGAGCAAGACTCCCATCTCAAAAAAAA	65254
Qy	2144	CACACTAAAAATTAATAAAGAGAGGAGATAGTGGGTGTGGTGCACACCTGCAAT	2203
Db	65255	AAAA--AAAAAANAANAANAANAANAAGCCAGGCACAATGGCTCAGCCCTGTAT	65312
Qy	2204	CCCAGCACTTTGGAAGGCCAGAGTGGGCGAGATCATCTGAGGCCAGGAGTCTCAAGACCAGC	2263
Db	65313	CTCAGCACTTTGGAGGCTGAGGCAGCGGATCACTGAGGTGGGAGTTTCGAGACCAG-	65372
Qy	2264	CTGGCTAACATGGTGAATCCTATCTCTACAAAAAATACAAAAATACAGGCGTGGTG	2323
Db	65372	CTGACCAACATAAAGAAACCCCGTCTCTACTAAAAAATACAGAAATTTAGCCGAGTATGTG	65431
Qy	2324	GTGGGCACCTGTACTGGGAGGTTGCCACCCAGCACTACTGGGAGGCTCAGTCAAGAGAT	2383
Db	65432	CGCATCGCTGTAT------CCACAGTACTCGGAGGCTGAGCAGGAGAT	65478
Qy	2384	CGCTTGAACCTGGGAGGCGAGGTGGCGTCAGCTGAGATGGTGCACCTGCACCTCCAGCC	2443
Db	65479	CGCTTGAACCTGGGAGGCGAGGCTGTGATGCGCGAGATTTGGCCATTGCATCCAGCC	65538
Qy	2444	TGGCGAAGAGGCACTCTGTCTCCAAAAAAGAGAGGAGGAGACACAGAGACACA	2503
Db	65539	TGGCAACAGAGCGAAACTCTGTTTCAAAAAACAAAAACAAAAAATAAATAATCGTCT	65598
Qy	2504	CAGAGAGAAAG 2515 	
Db	65599	CAGAGAAAAAGAG 65610	
RESULT	14		
LOCUS	AC005372/c		
DEFINITION	Homo sapiens chromosome 5, p1 clone 1195e2 (LBNL H73), complete sequence.	DNA	PRI 01-AUG-1998
ACCESSION	AC005372		
NID	G3367502		
VERSION	AC005372.1		
KEYWORDS	HTG.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
AUTHORS	Kimberly W., Bondoc, M., Cheng, J., Connolly, K.S., Gunning, K.M., Kadner, K., Miguei, T., Miller, C., Ptiluck, S., Pollard, M., Rojeski, H., Subramanian, S. and Martin, C.H.		
TITLE	Sequencing of human chromosome 5		
JOURNAL	Unpublished		

ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominoidea; Homo.
1 (bases 1 to 44957)
Kimberly, W., Bondoc, M., Cheng, J., Connolly, K. S., Gunning, K. M.,
Kadner K., Maguei, T., Miller, C., Pittluck, S., Pollard, M.,
Rojas, H., Subramanian, S. and Martin, C. H.
Sequencing of human chromosome 5
Unpublished

```

REFERENCE          2 (bases 1 to 44957)
AUTHORS            Ricker,D.O.
TITLE              Large Scale Sequence Analysis and Annotation with the Sequence
                   Comparison Analysis (SCAN) System
JOURNAL            Unpublished
REFERENCE          3 (bases 1 to 44957)
AUTHORS            Kimmerly,W., Bondoc,M., Cheng,J., Connolly,K.S., Gunning,K.M.,
                   Davis,C.A., Kadner,K., Miguel,T., Pitluck,S., Pollard,M.,
                   Rojeski,H., Subramanian,S. and Martin,C.H.
TITLE              Direct Submission
JOURNAL            Submitted (01-AUG-1998) Human Genome Center, DOE Joint Genome
                   Institute, Lawrence Berkeley National Laboratory, MS 74-157,
                   Berkeley, CA 94720, U.S.A.
COMMENT            Sequence submitted by:
FEATURES           Location/Qualifiers
                   1..44957
                   /organism="Homo sapiens"
                   /db_xref="taxon:9606"
                   /map="5q"
                   /clone="1195e2"
                   /chromosome="5"
                   /note="LBNL H73"
                   212..680
                   /rpt_family="Alu"
                   /complement(1329..1587)
                   /rpt_family="Alu"
                   1625..1921
                   /rpt_family="Alu"
                   /complement(2061..2351)
                   /rpt_family="Alu"
                   2741..3016
                   /rpt_family="Alu"
                   /rpt_family="Alu"
                   3279..3302
                   /note="(T)24"
                   /rpt_type=tandem
                   /rpt_unit=T
                   /rpt_unit=T
                   /complement(3285..3903)
                   /rpt_family="Alu"
                   /complement(4010..4138)
                   /rpt_family="Alu"
                   4153..4176
                   /note="(T)24"
                   /rpt_type=tandem
                   /rpt_unit=T
                   /complement(4159..4437)
                   /rpt_family="Alu"
                   4483..4785
                   /rpt_family="Alu"
                   join(4854..4952,7649..7813,11526..11678,15096..15209,
                   15884..16048,16908..17030,20748..20873,23801..23884,
                   25757..25846,30119..30313,31286..31450)
                   /standard_name="RNA helicase (Rat homolog)"
                   /note="89-100% protein identity PIR:A57514"
                   4856..4956
                   /note="GRAIL 2 excellent exon, frame 2"
                   /complement(5104..5336)
                   /rpt_family="Alu"
                   5373..5659
                   /rpt_family="Alu"
                   5913..6482
                   /rpt_family="Alu"
                   /complement(6552..6749)
                   /rpt_family="Alu"
                   /complement(6765..7042)
                   /rpt_family="Alu"
                   7649..7813
                   /note="GRAIL 2 excellent exon, frame 1"
                   /complement(8693..8801)
                   /rpt_family="Alu"
                   /complement(8811..9195)
                   /rpt_family="Alu"
                   /complement(9385..9732)

repeat_region     /rpt_family="Alu"
                   complement(9842..10511)
                   /rpt_family="Tigter2"
                   10067..10357
                   /rpt_family="Alu"
                   /complement(10526..10569)
                   /rpt_family="Tigter2"
                   10526..10569
                   /rpt_family="MER8"
                   10599..11178
                   /rpt_family="Alu"
                   11527..11678
                   /note="GRAIL 2 excellent exon, frame 2"
                   /complement(11974..12260)
                   /rpt_family="Alu"
                   /complement(12535..12810)
                   /rpt_family="Alu"
                   /complement(12860..13124)
                   /rpt_family="Alu"
                   /complement(13152..13436)
                   /rpt_family="Alu"
                   /complement(13489..13719)
                   /rpt_family="Alu"
                   13780..14047
                   /rpt_family="Alu"
                   14053..14098
                   /note="(GT)23"
                   /rpt_type=tandem
                   /rpt_unit=GT
                   14256..14500
                   /rpt_family="Alu"
                   14524..14563
                   /note="(AATA)10"
                   /rpt_type=tandem
                   /rpt_unit=AATA
                   14723..14999
                   /rpt_family="Alu"
                   15096..15297
                   /note="GRAIL 2 excellent exon, frame 2"
                   15884..16049
                   /note="GRAIL 2 excellent exon, frame 1"
                   16195..16467
                   /rpt_family="Alu"
                   /complement(16478..16798)
                   /rpt_family="Alu"
                   16909..17029
                   /note="GRAIL 2 excellent exon, frame 2"
                   /complement(17201..17504)
                   /rpt_family="Alu"
                   /complement(17533..17790)
                   /rpt_family="Alu"
                   17829..18122
                   /rpt_family="Alu"
                   /complement(18147..18462)
                   /rpt_family="Alu"
                   /complement(18520..19276)
                   /rpt_family="Alu"
                   20757..20870
                   /note="GRAIL 2 excellent exon, frame 2"
                   /complement(21133..21423)
                   /rpt_family="Alu"
                   21856..22123
                   /rpt_family="Alu"
                   22185..22208
                   /note="(GT)8"
                   /rpt_type=tandem
                   /rpt_unit=GT
                   /complement(22254..22429)
                   /rpt_family="Alu"
                   23024..23330
                   /rpt_family="Alu"
                   23801..23879
                   /note="GRAIL 2 excellent exon, frame 1"

```


This sequence was generated from part of bacterial clone contigs of human chromosome X, constructed by the Sanger Centre Chromosome X Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/ChrX> 426N21 is from the library RGC13 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong. For further details see <http://bacpac.med.buffalo.edu/VBC01r>.

```

FEATURES
source
Location/Qualifiers
1. .44826
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="X"
/clone="426N21"
/clone_lib="RPC13"
39. .350
/note="AlusX repeat: matches 1. .312 of consensus"
<1386. .>1723
/note="match: GSS AQ131901"
2516. .2557
/note="MER5A repeat: matches 146. .189 of consensus"
2556. .2641
/note="MIR repeat: matches 28. .116 of consensus"
2642. .2944
/note="AluY repeat: matches 4. .297 of consensus"
2954. .3288
/note="AlusX repeat: matches 1. .312 of consensus"
3269. .3282
/note="MIR repeat: matches 124. .137 of consensus"
3283. .3587
/note="AlusX repeat: matches 1. .305 of consensus"
3588. .3720
/note="MIR repeat: matches 137. .262 of consensus"
3732. .4032
/note="AluY repeat: matches 1. .301 of consensus"
4425. .4605
/note="Alusg/x repeat: matches 131. .311 of consensus"
4606. .4921
/note="Alusq repeat: matches 1. .313 of consensus"
4922. .4958
/note="Alusg/x repeat: matches 95. .131 of consensus"
4961. .5008
/note="Alusg/x repeat: matches 129. .185 of consensus"
5009. .5142
/note="AluJo/FLAM repeat: matches 2. .185 of consensus"
5151. .5464
/note="Alusg repeat: matches 2. .307 of consensus"
5467. .5656
/note="AluJb repeat: matches 119. .305 of consensus"
5657. .5969
/note="AlusX repeat: matches 1. .299 of consensus"
6219. .6357
/note="AluJb repeat: matches 1. .141 of consensus"
6906. .7204
/note="AlusX repeat: matches 1. .298 of consensus"
7288. .7587
/note="Alusq repeat: matches 1. .309 of consensus"
7595. .7694
/note="MIR repeat: matches 107. .223 of consensus"
8509. .8807
/note="AluYb8 repeat: matches 1. .305 of consensus"
8939. .9241
/note="AlusX repeat: matches 1. .304 of consensus"
10115. .10418
/note="AluJo repeat: matches 1. .306 of consensus"
11014. .11319
/note="AlusX repeat: matches 1. .308 of consensus"
11483. .11512
/note="15 copies 2 mer at 87% conserved"
12044. .12121
/note="MER5A repeat: matches 85. .162 of consensus"
12548. .12577

```

[illegible]

```

/NOTE="15 copies 2 mer tt 87% conserved"
12800. .13093
/NOTE="AluJb repeat: matches 17. .311 of consensus"
13197. .13475
/NOTE="AluXg repeat: matches 1. .281 of consensus"
13548. .13839
/NOTE="AluSg repeat: matches 1. .300 of consensus"
13979. .14371
/NOTE="L1R37A repeat: matches 7. .426 of consensus"
14397. .14440
/NOTE="22 copies 2 mer tg 89% conserved"
14498. .14579
/NOTE="L2 repeat: matches 2635. .2710 of consensus"
14580. .14890
/NOTE="AluSp repeat: matches 1. .313 of consensus"
14891. .16029
/NOTE="L2 repeat: matches 1456. .2635 of consensus"
16038. .16091
/NOTE="18 copies 3 mer gag 74% conserved"
16570. .16861
/NOTE="L2 repeat: matches 2288. .2517 of consensus"
16862. .17172
/NOTE="AluSp repeat: matches 1. .313 of consensus"
17173. .17369
/NOTE="L2 repeat: matches 2517. .2731 of consensus"
17744. .18046
/NOTE="AluXg repeat: matches 1. .303 of consensus"
19119. .19431
/NOTE="AluXg repeat: matches 1. .299 of consensus"
19861. .20160
/NOTE="AluJ repeat: matches 1. .301 of consensus"
20205. .20371
/NOTE="AluJo repeat: matches 1. .169 of consensus"
21016. .21331
/NOTE="AluXg repeat: matches 1. .312 of consensus"
21318. .21413
/NOTE="Single clone region. single clone in alu"
21429. .21721
/NOTE="AluSg repeat: matches 1. .295 of consensus"
21767. .22064
/NOTE="AluSp repeat: matches 1. .299 of consensus"
22570. .22639
/NOTE="L2 repeat: matches 2430. .2498 of consensus"
22685. .22984
/NOTE="AluSp repeat: matches 1. .308 of consensus"
24345. .24535
/NOTE="AluSg/x repeat: matches 128. .312 of consensus"
24729. .25043
/NOTE="AluXg repeat: matches 1. .311 of consensus"
complement(26632. .>26963)
/NOTE="match: 3' EST AI087196 clone IMAGE:1679850"
26893. .27480
/NOTE="14 copies 42 mer 59% conserved"
27062. .27493
/NOTE="9 copies 48 mer 63% conserved"
27305. .27516
/NOTE="106 copies 2 mer cc 55% conserved"
30587. .30660
/NOTE="Charlie4 repeat: matches 1882. .1956 of consensus"
30935. .31131
/NOTE="MIR repeat: matches 1. .206 of consensus"
31143. .31437
/NOTE="AluSg repeat: matches 1. .299 of consensus"
31553. .31744
/NOTE="MIR repeat: matches 7. .162 of consensus"
31745. .32000
/NOTE="L1M87 repeat: matches 5923. .6173 of consensus"
32001. .32089
/NOTE="MIR repeat: matches 162. .259 of consensus"
32690. .32767
/NOTE="AluJ repeat: matches 1. .78 of consensus"
33718. .33857
/NOTE="MIR repeat: matches 82. .216 of consensus"

```


This Page Blank (uspto)

**This Page is Inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

☐ BLACK BORDERS

☒ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES

☐ FADED TEXT OR DRAWING

☐ BLURRED OR ILLEGIBLE TEXT OR DRAWING

☐ SKEWED/SLANTED IMAGES

☐ COLOR OR BLACK AND WHITE PHOTOGRAPHS

☐ GRAY SCALE DOCUMENTS

☐ LINES OR MARKS ON ORIGINAL DOCUMENT

☒ REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY

☐ OTHER: _____

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.

This Page Blank (uspto)